



**Carlos Miguel
Aidos Almeida**

**Sistema de Autenticação pessoal baseado em sinais
EEG**



**Carlos Miguel
Aidos Almeida**

Sistema de Autenticação pessoal baseado em sinais de EEG

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica da Dra. Pétia Georgieva, e da Dra. Ana Maria Tomé Professoras do Departamento de Electrónica Telecomunicações e Informática da Universidade de Aveiro

Dedico este trabalho aos familiares e amigos e me ajudaram e apoiaram nesta etapa da minha vida.

o júri

Presidente

Prof. Doutor José Fernando Rocha Pereira
Professor Associado com Agregação da Universidade de Aveiro

vogais

Professora Doutora Ana Maria Perfeito Tomé (Co-orientadora)
Professora Associada da Universidade de Aveiro

Professor Doutor Luís Miguel Almeida da Silva
Professor Auxiliar do Departamento de Matemática para Ciência e Tecnologia
da Escola de Ciências da Universidade do Minho

Professora Doutora Pétia Georgieva Georgieva
Professora Auxiliar da Universidade de Aveiro (Orientadora)

Agradecimentos

É com muito gosto que aproveito para agradecer a todos os que me ajudaram durante este ano de trabalho.

Gostaria de agradecer às orientadoras Prof. Dra. Pétia Georgieva e Prof. Dra. Ana Maria Tomé que mostraram sempre a maior disponibilidade para me ensinar e ajudar em diversas questões.

Queria também agradecer à minha família pelo apoio moral e financeiro que sempre me deram e também a todos os meus amigos que sempre me acompanharam ao longo deste percurso.

palavras-chave

Autenticação, Biometria, Tecnologias Biométricas, Electroencefalograma, Bandas de Frequências, Potenciais Evocados Visuais, Filtragem, Knn, Classificação, Pós Processamento

Resumo

Neste trabalho pretende-se estudar a actividade cerebral resultante de estímulos visuais, que é medida através de Electroencefalograma (EEG). O objectivo é verificar se, através da análise dos sinais EEG, será possível implementar um sistema de autenticação de indivíduos on-line.

Este projecto encontra-se dividido em 5 capítulos. O primeiro consiste numa introdução ao conceito da biometria. No segundo, Estado da Arte, é exposta um pouco da história que serve de base à biometria actual, são referenciadas algumas das tecnologias biométricas mais proeminentes dos dias de hoje, são explicados alguns conceitos necessários para entender como funciona a técnica do EEG e, por último, é feita uma pequena introdução ao classificador utilizado (KNN). O terceiro capítulo, que tem por nome Base de Dados e Métodos, é o capítulo onde é retratada a forma como se obtiveram os dados que serviram de base ao projecto e, depois, a forma como estes foram processados. No quarto capítulo (Classificação e estudos Efectuados) são apresentados todos os testes e estudos realizados de forma a otimizar o processo de classificação e também os resultados que daí se obtiveram. No último capítulo são feitas as conclusões finais sobre o trabalho e os possíveis trabalhos, de continuação, a implementar futuramente.

keywords

Authentication, Biometrics, Biometrics Technologys, Electroencephalogram, Frequency Bands, Visual Evoked Potencials, Filtering, Knn, Classification, Post Processing

Abstract

This paper aims to study the brain activity resulting from visual stimuli, which is measured by electroencephalograms (EEG). The gold is to verify, through EEG signals analysis, if it will be possible to implement an online authentication system of individuals.

This project is divided into 5 chapters. The first one is an introduction to the concept of biometrics. In the second, State of the Art, is set a little of the story that serves as the basis for the current biometrics techniques, are listed some of the most prominent biometric technologies of today, are explained some concepts needed to understand how the EEG technique works and, finally, there is also a short introduction to the used classifier (KNN). The third chapter, which is called by Database and Methods, is the chapter that explains how the data was gathered and then, how it was processed. In the fourth chapter (Classification and studies) are presented all the tests and studies that were made in order to optimize the classifier, and also the results that were obtained from there. In the last chapter final conclusions are made about the work and the possible future work in this subject.

Índice

Capítulo 1 - Introdução	14
Capítulo 2 - Estado da Arte	16
2.1 - História	17
2.2 - Autenticação Biométrica Vs Identificação Biométrica	19
2.3 - Tecnologias Biométricas	20
2.3.1 - Impressão Digital	21
2.3.2 - Reconhecimento Facial	22
2.3.3 - Varrimento da Retina	23
2.3.4 – Reconhecimento pela Íris	24
2.3.5 – Reconhecimento de Voz	25
2.3.6 – Geometria da mão	25
2.3.7 – Reconhecimento das Veias	26
2.4 - Electroencefalograma (EEG)	27
2.5 - Potenciais Evocados	29
2.5.1 – Potenciais Evocados Visuais	30
2.6 - Classificadores	30
2.6.1 - K Nearest Neighbors (KNN)	30
Capítulo 3 - Bases de Dados e Métodos	33
3.1 - Base de Dados	34
3.1.1 – Resumo da Experiência	34
3.1.2 – Extracção dos Resultados da Experiência	35
3.2 - Processamento do sinal EEG	36
3.2.1 – Remoção das Épocas com Artefactos	38
3.2.2 – Filtragem do Sinal na Banda de Frequências Gama (30-50) Hz	39
3.2.3 – Cálculo e Normalização da Energia para cada Época	40
3.2.4 – Formação e Análise dos Vectors das Características	41

Capítulo 4 - Classificação e Estudos Efectuados	45
4.1 - Classificação	46
4.1.1 – Casos de Estudo	46
4.2 - Pós Processamento (P.P.) da saída do classificador	47
4.3 - Estudos Efectuados	51
4.3.1 - Treino e Validação com Tipos de Estímulos Distintos	51
4.3.2 - Treino e Validação com Tipos de Estímulos Semelhantes	52
4.3.3 - Variação da Ordem do Filtro	53
4.3.4 - Variação da Banda de Frequências Extraída	55
4.3.5 - Redução Progressiva do Número de Canais Utilizados	56
4.3.6 - Redução Progressiva do Número de Épocas Utilizados	59
4.3.7 - Alteração do Valor de K no Classificador	61
4.4 – Outros Pontos a Considerar	64
4.4.1 - Redução Progressiva do Número de Pessoas Utilizadas para Treino e Teste do Classificador	64
4.4.2 - Erro de Classificação Individual	66
Capítulo 5 - Conclusões	69
Bibliografia	73
Anexos	75

Índice (Figuras)

Fig.1 - Autenticação Vs Identificação	19
Fig.2 - Fases em comum entre as tecnologias biométricas para a extracção das características biométricas	20
Fig.3 - Autenticação por Impressão Digital	21
Fig.4 - Autenticação por Leitura da Íris	24
Fig.5 - Autenticação por Geometria da Mão	26
Fig.6 - Electroencefalograma (EEG)	27
Fig.7 - Bandas de Frequência no Sinal EEG	29
Fig.8 - Exemplo de Imagens utilizadas	34
Fig.9 - Período de duração dos Estímulos	35
Fig.10 - Localização dos Canais	35
Fig.11 - Fases do Processamento do Sinal EEG	37
Fig.12 - Potencial Eléctrico ao longo de 5 Épocas	38
Fig.13 - Potencial Eléctrico ao Longo de 5 Épocas (já com a segunda removida)	39
Fig.14 - Resposta em Magnitude e Fase do Filtro Butterworth de Ordem 10	40
Fig.15 - Vector de Características da Pessoa 1 (referente à primeira Época)	42
Fig.16 - Vectores de Características referentes à Pessoa 1	43
Fig.17 - Vectores de Características referentes à Pessoa 2	43
Fig.18 - Vectores de Características referentes à Pessoa 3	44
Fig.19 - Evolução do Erro ao longo do P.P. completo	49
Fig.20 - Treino e Validação com tipos de Estímulos distintos	52
Fig.21 - Treino e Validação com tipos de Estímulos semelhantes	52
Fig.22 - Variação da Ordem do Filtro (Ordem 6)	53
Fig.23 - Variação da Ordem do Filtro (Ordem 8)	54
Fig.24 - Variação da Banda da Ordem do Filtro	54
Fig.25 - Resultados para a Banda de Frequências Beta	55

Fig.26 - Variação da Banda de Frequências extraída	56
Fig.27 - Resultados para a utilização de apenas 15 Canais activos	57
Fig.28 - Resultados para a utilização de apenas 10 Canais activos	57
Fig.29 - Resultados para a utilização de apenas 3 Canais activos	58
Fig.30 - Redução progressiva do número de Canais utilizados	58
Fig.31 - Erro associado à Classificação	59
Fig.32 - Redução progressiva do número de Épocas utilizadas (utilização de 120, de 80 e de 40 Épocas)	60
Fig.33 - Erro associado à Classificação (K=3 Vs K=1)	61
Fig.34 - Erro associado à Classificação (K=3 Vs K=5)	62
Fig.35 - Erro associado à Classificação (K=3 Vs K=7)	62
Fig.36 - Erro associado à Classificação (K=3 Vs K=9)	62
Fig.37 - Erro associado à Classificação (K=3 Vs K=11)	63
Fig.38 - Erro associado à Classificação (K=3 Vs K=13)	63
Fig.39 - Erro associado à Classificação (K=3 Vs K=15)	63
Fig.40 - Resultados para a utilização de 13, de 11, de 8 e de 6 Pessoas tanto para Treino como para Teste do Classificador	65
Fig.41 - Redução progressiva do número de Pessoas utilizadas para Treino e Teste do Classificador	65
Fig.42 - Exemplo escolhido para apresentar o Erro de Classificação Individual	66
Fig.43 - Erro de Classificação Individual (primeiras 6 Pessoas)	66
Fig.44 - Erro de Classificação Individual (7 Pessoas seguintes)	67

Índice (Tabelas)

Tabela 1 - Explicação do Algoritmo KNN (1) -----	31
Tabela 2 - Explicação do Algoritmo KNN (2) -----	32
Tabela 3 - Valores arredondados das Energias por Canal (Primeira Época) -----	41
Tabela 4 - Valores arredondados das Energias Normalizadas por Canal (Primeira Época) -----	41
Tabela 5 - Parte da sequência temporal de valores à saída do classificador (ypred) ----	47
Tabela 6 - Pós Processamento (primeira Iteração) -----	48
Tabela 7 - Pós Processamento (segunda Iteração) -----	48
Tabela 8 - Pós Processamento (terceira Iteração) -----	49
Tabela 9 - Parte da sequência temporal de valores à saída do classificador (ypred) ----	50
Tabela 10 - Parte da sequência temporal de valores resultante do Pós Processamento completo (ypred5) -----	50
Tabela 11 - Erro de Classificação Individual sem PP -----	66
Tabela 12 - Erro de Classificação Individual com PP -----	66

Índice (Equações)

Equação 1 – Distância Euclidiana entre dois Pontos ----- 30

Equação 2 – Cálculo das Energias ----- 40

Capítulo 1

Introdução

A Biometria (do grego Bios = vida, metron = medida) consiste no uso de características biológicas em mecanismos de identificação/autenticação. Ao longo da última década, várias tecnologias biométricas têm sido exploradas e implementadas com sucesso na área da identificação e autenticação pessoal.

Para que uma característica humana (física ou comportamental) possa ser utilizada para identificação/autenticação pessoal, deve satisfazer os seguintes requisitos [R1]:

- **Universalidade:** deve existir em todas as pessoas
- **Singularidade:** deve ser distinta em cada pessoa
- **Permanência:** não pode variar com o tempo

As tecnologias biométricas são, normalmente, classificadas como comportamentais ou físicas, de acordo com a classificação das características avaliadas. Porém, podem também ser classificadas como colaborativas (utilizador tem conhecimento da sua existência e participa conscientemente no processo) ou furtivas (utilizadas sem o conhecimento daquele que é identificado ou autenticado) [R2].

A autenticação de uma pessoa é a base da segurança da informação, e pode ser definida como um processo de verificação, ou seja, verificar se uma pessoa é realmente quem ela diz ser.

Na área da segurança, os métodos normalmente mais utilizados para a autenticação pessoal, são os métodos baseados no que a pessoa sabe (no caso das senhas), ou que possui (no caso de cartões ou chaves). Estes métodos têm uma vasta aplicação em vários domínios sendo, no entanto, vulneráveis a fraudes (senhas podem ser esquecidas e cartões podem ser roubados ou perdidos). É, neste cenário, que surgem os métodos baseados na biometria humana. Estes, são de reconhecimento automático de um indivíduo, a partir das características físicas ou de comportamento (baseadas no que o indivíduo é ou faz).

Através destes métodos, como a impressão digital, a íris, a voz ou a própria forma da face, existe uma maior dificuldade para que alguém consiga realizar uma autenticação enganosa fazendo-se passar por outra pessoa.

Porém, nenhuma das características biométricas mencionadas anteriormente, impõe que o sujeito autenticado responda a estímulos, o que pressupõe que eventualmente possam existir falsificações destas características. Daí, surgir a necessidade de autenticar indivíduos a partir das suas respostas cerebrais a estímulos exteriores, admitindo que os estímulos, mas não as suas respostas, podem variar entre diferentes ocasiões em que é feita a operação de autenticação.

Este trabalho pretende estudar a actividade cerebral resultante de estímulos visuais, que é medida através de Electroencefalograma (EEG). O objectivo é verificar se, através da análise dos sinais EEG será possível implementar um sistema de autenticação de indivíduos on-line.

Capitulo 2

Estado da Arte

2.1 História

De um modo pouco sofisticado pode dizer-se que a biometria já existe há séculos. Partes do nosso corpo e aspectos do nosso comportamento têm sido usados, ao longo dos tempos, como um modo de identificação, visto que, no nosso dia-a-dia, identificamos e lembramo-nos das pessoas pelo som da sua voz, pela textura do seu rosto, ou por alguma outra característica física ou comportamental.

Francis Galton é um dos fundadores da biometria como ciência. A sua pesquisa em habilidades e disposições mentais foi pioneira em demonstrar que vários traços são genéticos. A paixão de Galton pela medição permitiu que ele abrisse o Laboratório de Antropométrica na Exposição Internacional de Saúde em 1884, onde recolheu estatísticas de milhares de pessoas.

Em 1892, Galton inventou o primeiro sistema moderno de impressão digital, usado por vários departamentos de polícia em todo o mundo. A impressão digital foi durante vários anos a forma mais confiável de identificação até surgir mais tarde no século XX a identificação por ADN (ácido desoxirribonucleico).

Os avanços comerciais na área da biometria começaram na década de setenta. Durante este período, um sistema chamado *Identimat* foi criado e instalado em vários locais para controlo de acesso. Este sistema media a forma da mão, dando principal importância ao tamanho dos dedos do utilizador.

A produção do *Identimat* teve fim nos anos oitenta. No entanto, o seu uso foi pioneiro na aplicação da geometria da mão como tecnologia biométrica.

Paralelamente ao desenvolvimento da geometria da mão como tecnologia biométrica, nas décadas de sessenta e setenta, a biometria digital teve bastantes progressos. No decorrer destes anos, algumas companhias estavam envolvidas com a identificação automática das imagens digitais como forma de auxílio às forças policiais.

No final dos anos sessenta, o FBI (Federal Bureau of Investigation) começou a fazer a análise das imagens digitais automaticamente, e na metade da década de setenta já havia instalado uma grande quantidade de sistemas de scanners digitais automáticos. Desde então, o papel da biometria no que refere à segurança tem aumentado e, os Automated Fingerprint Identification Systems (AFIS) são utilizados por um número significativo de forças policiais em todo mundo [R3].

Outras técnicas têm evoluído ao lado das biometrias pioneiras dos anos sessenta e setenta. O primeiro sistema a analisar o padrão único da retina foi introduzido na metade dos anos oitenta. Enquanto isso, o trabalho do Dr. John Daugman da Universidade de Cambridge abriu o caminho para o desenvolvimento da tecnologia da íris. A actual verificação de voz possui raízes assentadas nos empreendimentos tecnológicos dos anos setenta enquanto biometrias como a verificação de assinaturas e reconhecimento facial eram relativamente novas na indústria.

A investigação na área da biometria e o desenvolvimento de sistemas comerciais biométricos têm vindo a aumentar. Os avanços científicos e tecnológicos e o grande interesse por parte das empresas são essenciais para se redefinir o desempenho das tecnologias biométricas já existentes e para se criarem novas e mais inovadoras técnicas.

Porém:

Como referido anteriormente, com a proliferação das tecnologias informáticas e o avanço dos estudos sobre biometria, tornou-se viável a implementação de autenticação por recurso a características biométricas de indivíduos. No entanto, estas soluções, além das dificuldades técnicas, acarretavam certas dificuldades de nível social.

Países como a Austrália, Canadá, Estados Unidos e Nova Zelândia testemunharam uma inquietação pública quanto aos esquemas de identificação. Entre os vários receios citados incluem-se [R4]:

- Que as pessoas sejam desumanizadas ao serem reduzidas a códigos
- Que o sistema potencie o poder sobre os indivíduos por parte de determinadas organizações e até do estado
- Que a identificação de alta integração envolva a inversão da apropriada relação entre o cidadão e estado
- Que o sistema seja conduzido por uma burocracia tecnologicamente assistida, ao invés de governos eleitos
- Que isenções e excepções existam para organizações e indivíduos poderosos
- Que estes esquemas de identificação sejam os mecanismos previstos em profecias religiosas como, por exemplo, *a Marca da Besta*

Com a generalização de equipamentos de características biométricas e com a sua divulgação (por exemplo em filmes de sucesso), o cidadão comum encara hoje a autenticação biométrica como algo que lhe é familiar.

2.2 Autenticação Biométrica Vs Identificação Biométrica

Recorrendo a conceitos mais gerais, Autenticação refere-se ao problema de confirmar ou negar uma alegada identidade de um indivíduo, enquanto Identificação refere-se ao problema de estabelecer a identidade, desconhecida à partida, de um indivíduo [R5].

O processo de Identificação pode ser visto como um processo de reconhecimento, numa espécie de pesquisa aberta. Neste processo, inicialmente, são apresentados ao sistema os dados do indivíduo a identificar e, posteriormente, estes dados vão ser comparados com todos os templates da base de dados do sistema. Este processo também é conhecido como comparação um-para-muitos (1:N).

O processo de Autenticação é visto mais como um processo de verificação numa espécie de pesquisa fechada. Neste caso, os dados do indivíduo a autenticar são comparados com os dados desse mesmo indivíduo, dados esses já existentes na base de dados. Em caso de verificação positiva o sistema biométrico confirma que o solicitante é, de facto, quem afirma ser.

A chave para a distinção entre Identificação e Autenticação reside nas perguntas efectuadas ao sistema:

- **Identificação:** Quem é esta pessoa?
- **Autenticação:** Esta pessoa é quem ela afirma ser?

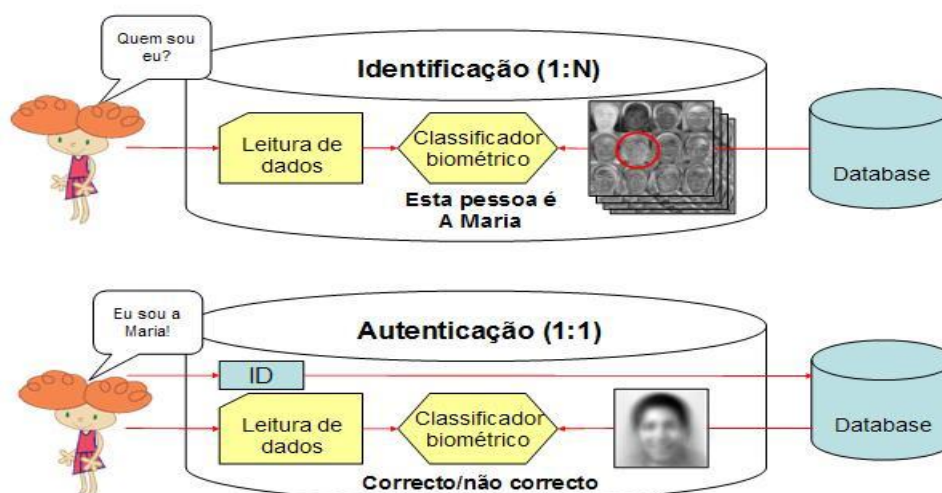


Fig.1 -> Autenticação Vs Identificação

2.3 Tecnologias Biométricas

Como referido anteriormente, existem várias tecnologias biométricas que são constantemente utilizadas, para questões de segurança, nos dias de hoje. Para cada uma destas tecnologias, podem existir várias formas distintas de realizar o processo de extracção das características biométricas.

Estas formas, embora distintas como referido, têm em comum as fases necessárias para a extracção das tais características biométricas. Estas fases podem ser visualizadas através da figura seguinte (fig.2):

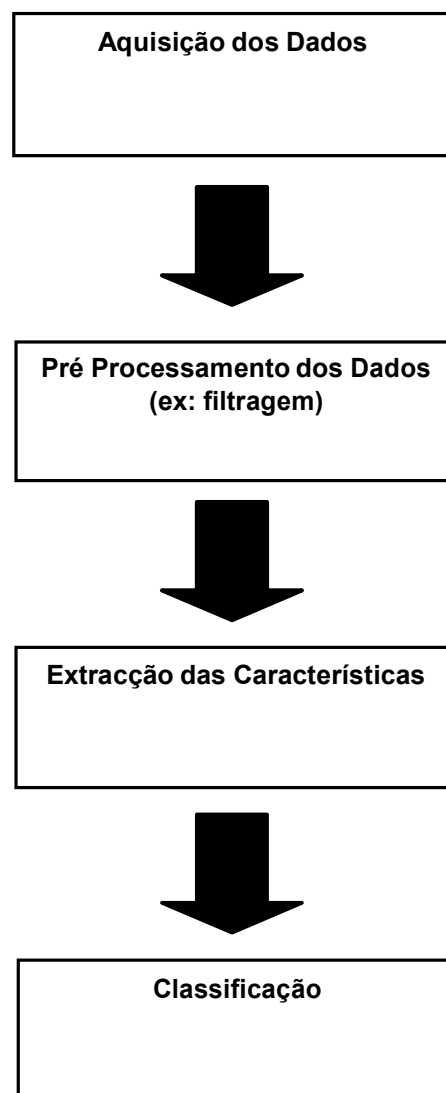


Fig.2 -> Fases em comum entre as tecnologias biométricas para a extracção das características biométricas

Seguidamente será retratada, para cada tecnologia biométrica, uma forma de obter as características biométricas que se lhe adequam.

2.3.1 Impressão Digital

A tecnologia de impressão digital é a tecnologia de identificação/autenticação biométrica mais proeminente, sendo utilizada por milhões de pessoas em todo o mundo.

A impressão digital é composta por vários sulcos que, na sua formação, apresentam diferenças chamadas de pontos de minúcias (pontos característicos) [R6]. Cada um desses pontos apresenta características únicas, que podem ser medidas através do sensor biométrico. Este sensor é um dispositivo capaz de efectuar o registo da impressão digital com um elevado grau de precisão, de forma a registar os traços (minúcias) que a definem.

Das imagens (templates) capturadas, são extraídos (através de um algoritmo específico) pontos particulares da imagem. Esta informação é depois convertida num único modelo matemático comparável a uma palavra passe de 60 dígitos. Neste processo não é guardada nenhuma imagem concreta da impressão digital.

Esta tecnologia tem tido várias aplicações, nomeadamente ao nível do controlo de acessos a áreas restritas, protecção de equipamentos valiosos, autenticação de pais em jardins-de-infância, identificação de pessoas ao nível criminal entre outras.

No entanto, esta tecnologia apresenta uma desvantagem, visto que, está estimado que 1 a 2% da população não possa usar este tipo de tecnologia [R7]. Isto deve-se ao facto de que a profissão de algumas pessoas reduz a qualidade das suas impressões digitais (trabalhos manuais pesados, trabalhos que envolvam a utilização de químicos, entre outras).



Fig.3 -> Autenticação por Impressão Digital

2.3.2 Reconhecimento Facial

A tecnologia de reconhecimento facial não difere muito da capacidade que o Homem tem de reconhecer a identidade dos outros através da face [R6]. Fazendo uso de software e algoritmos específicos, um computador localiza faces humanas num dado campo visual, através de câmaras de vigilância, fazendo depois a comparação com uma base de dados previamente construída.

Para reconhecer o rosto de uma pessoa, os programas mapeiam a geometria e as proporções da face. São registados vários pontos delimitadores na face, os quais permitem definir proporções, distâncias e a forma de cada elemento do rosto. Com base nesses dados, são iniciadas as comparações. Os pontos principais desta análise são os olhos, o nariz, as maçãs do rosto, as orelhas e lábios.

Nesta tecnologia a câmara captura uma fotografia do rosto humano, que é mapeada numa série de 128 números, conhecidos como coeficientes. Estes são depois processados de forma a compor um arranjo único e bidimensional da disposição de áreas claras e escuras do rosto. Distâncias, tamanhos e a forma de cada traço são então registados.

A tecnologia de reconhecimento facial assume que as medidas do rosto nunca se alteram mesmo que a pessoa deixe crescer a barba, passe a usar óculos ou, por ventura, seja submetida a uma operação plástica. Isto constitui uma limitação desta tecnologia embora possa ser ultrapassada com o desenvolvimento do reconhecimento facial em 3D.

O reconhecimento facial mesmo a duas dimensões apresenta uma taxa de sucesso interessante. Estes sistemas estão cada vez mais sofisticados, visto que exigem ao utilizador uma prova de que realmente estão a expor uma cara humana, como piscar os olhos ou sorrir, como forma de prevenir falsificações (ex: utilização de mascaras).

Vantagens:

- Tecnologia não intrusiva
- Podem ser criados templates faciais sem a presença física do indivíduo
- Vários países possuem bases de dados com fotografias de faces, fruto de processos já existentes, como o bilhete de identidade ou o próprio passaporte.

2.3.3 Varrimento da Retina

A retina é a camada das veias sanguíneas situada na parte de trás do olho. Assim, a retina, forma um padrão único, que apenas se altera após a morte.

Os sistemas biométricos baseados na leitura da retina analisam a camada de vasos sanguíneos situada na parte de trás do olho, através da utilização de uma fonte de luz de baixa intensidade para, opticamente, reconhecer padrões únicos. Esta tecnologia pode atingir elevados níveis de precisão, mas requer que o utilizador olhe para dentro de um receptáculo e foque um determinado ponto, o que não é conveniente para utilizadores que usem óculos ou que receiem o contacto próximo com o leitor [R8].

Este tipo de biometria é reconhecido como um dos métodos biométricos mais seguros, visto que utiliza um procedimento preciso de registo, que envolve o alinhamento da vista de forma a alcançar uma leitura optimizada.

O processo de varredura da retina ocular envolve três fases [R9].

Na primeira fase, Captura, o olho é posicionado em frente ao sistema, a uma certa distância de um leitor ocular, de modo a que o utilizador possa visualizar um ponto verde por alguns segundos. Em seguida, o leitor captura o padrão de retina do mesmo. Uma área conhecida como fóvea, situada no centro da retina, é analisada e, é então capturado um padrão único das veias sanguíneas.

Na segunda fase, Extracção, o equipamento biométrico vai mapear a posição das veias sanguíneas, de modo a extrair uma representação matemática que irá ser armazenada como template.

A última fase tem como nome Comparação. Esta fase tem como objectivo principal proceder à autenticação ou identificação do utilizador.

Vantagens:

- Uma das tecnologias biométricas mais seguras.
- Inalterável com o passar dos anos (estrutura não se altera).

Esta técnica apresenta, no entanto, uma desvantagem visto ser uma técnica demasiado invasiva já que o utilizador tem que olhar fixamente para um ponto de luz a uma curta distância. Como resultado, esta técnica tem impulsionado o caminho da utilização da análise da íris, que é também um método de alta confiabilidade, sendo, no entanto, menos invasivo.

2.3.4 Reconhecimento pela Íris

A íris é o anel colorido que circunda a pupila do olho. Cada íris possui uma estrutura única, caracterizando um padrão complexo. Pode resultar numa combinação de características específicas como cor, glândula, filamentos, sardas, sulcos radiais e estrias. Estas características são altamente complexas e únicas, visto que, a probabilidade de duas íris serem idênticas é estimada em cerca de 1 caso em 10^{78} [R6].

A ideia de utilizar o padrão da íris para identificação de pessoas foi originalmente proposto em 1936 pelo oftalmologista Franck Burch. Em 1987, dois oftalmologistas, Aran Safir e Leonard Flom patentearam esta ideia, e, em 1989, pediram a John Daugman que procurasse criar um padrão e algoritmo de modo a que tornasse possível o reconhecimento de indivíduos através da íris. Estes algoritmos, que Daugman patenteou em 1994, são detidos pela Iridan Technologies/Securimetrics e são a base de todos os actuais equipamentos e soluções de reconhecimento através da íris.

A tecnologia da íris depende de visão computadorizada, reconhecimento de padrões, inferência estatística e óptica. Esta tecnologia propõe-se a identificar/autenticar pessoas de forma extremamente segura e em tempo real, fazendo uso de análises matemáticas dos padrões que existem na íris. Como a íris é um órgão interno e protegido, cuja textura é estável ao longo dos tempos, permite que um único registo seja utilizável durante toda a vida.

Esta tecnologia, actualmente, é utilizada para substituição de passaportes, segurança aérea, controlo de acessos a áreas de alto risco, entre outras aplicações.

Vantagens:

- É a tecnologia mais robusta e viável existente no mercado
- Os templates uma vez guardados, não precisam de ser novamente adquiridos, visto que, a íris não se altera durante toda a vida utilizador.



Fig.4 -> Autenticação por Leitura da Íris

2.3.5 Reconhecimento de Voz

O som da voz humana é causado pela ressonância das cordas vocais. O comprimento da corda vocal, o formato da boca e as cavidades nasais têm um papel muito importante, visto que, o som é medido quando afectado por essas características específicas.

A técnica de medição da voz pode usar ou não métodos dependentes de texto. Actualmente, as técnicas dependentes de texto são dominantes nos sistemas comerciais disponíveis de identificação de fala.

A biometria de identificação/autenticação de fala é particularmente útil para aplicações baseadas na telefonia. As redes telefónicas são um dos meios mais utilizados para comunicação e os sistemas biométricos podem facilmente ser incorporados em redes telefónicas privadas ou públicas. Porém, barulhos e interferências no ambiente onde se encontra a rede telefónica podem afectar o desempenho dos sistemas de identificação da fala.

O processo de autenticação da fala passa por 3 fases [R9]. Na primeira fase, Captura, o utilizador, através dum microfone, dita uma frase previamente seleccionada (protocolo dependente do texto) ou aleatória (protocolo independente do texto). Este processo pode ser repetido várias vezes de forma a ser possível construir-se um perfil da voz.

Na segunda fase, Extracção, o equipamento biométrico extrai as características únicas da voz criando assim um template.

Na terceira e última fase, Comparação, a verificação um-para-um (1:1) é o método preferencial, ou seja, o utilizador fala num microfone e este novo exemplo da voz é comparado com o template armazenado.

Vantagens:

- É uma das tecnologias biométricas menos invasivas.
- Baixos Custos

2.3.6 Geometria da mão

O reconhecimento da geometria da mão resulta de uma análise das características da própria mão, como a sua forma, o comprimento dos dedos e as suas linhas características [R5]. Pode-se obter diferentes níveis de segurança neste sistema consoante se utilizem as características em si, a posição das características

relativamente a um ponto fixo ou a fixação de vários pontos e as distâncias das características relativamente a todos eles.

De realçar que a geometria da mão (tal como os algoritmos de hoje a interpretam) tem a vantagem de ser facilmente combinada com outras biometrias como, por exemplo, a impressão digital.

Torna-se quase impossível obter secretamente informações precisas sobre a geometria da mão de uma pessoa, a menos que haja cooperação por parte da pessoa em questão. Quanto à estabilidade, deve-se ressaltar que a geometria da mão muda de acordo com a idade e, ocasionalmente, pode também mudar com a perda ou ganho de peso [R2].

Esta é uma técnica ideal para ambientes onde o acesso a áreas restritas e movimentadas necessita de ser efectuada de uma forma rápida e segura, como no controlo de acesso de funcionários a uma empresa ou até mesmo em universidades.



Fig.5 -> Autenticação por Geometria da Mão

2.3.7 Reconhecimento das Veias

O reconhecimento através das veias é possível pois o padrão das mesmas varia de indivíduo para indivíduo. Recorrendo a luz infravermelha é possível fazer sobressair as características subcutâneas existentes em cada um de nós, em particular no que diz respeito às veias.

A hemoglobina presente no sangue absorve a luz de forma a tornar esta técnica possível e segura. O padrão das veias é então transformado em código binário, de forma a ser comparado futuramente e assim garantir a identificação/autenticação do indivíduo. Existem vários tipos de tecnologia de reconhecimento das veias, que podem ser dos dedos, da palma ou das costas da mão e até dos pulsos. O conceito de tratamento subjacente a cada uma destas técnicas é idêntico.

Esta tecnologia, não possuindo ainda um vasto leque de aplicações, tem vindo a ser aplicada em situações como, por exemplo, login em computadores. A segurança inerente a esta tecnologia fica patente quando entidades bancárias apostam na mesma para caixas multibanco.

Vantagens:

- Padrão estável mesmo com o crescimento das veias.
- Tecnologia higiénica pois não obriga a contacto com os equipamentos de recolha.

As tecnologias biométricas, acima referidas, são tecnologias passivas. Nesse sentido, uma falsificação é possível e mais provável do que através de técnicas que exijam resposta a estímulos. Daí a importância do desenvolvimento da tecnologia EEG no âmbito da autenticação de um indivíduo.

2.4 Electroencefalograma (EEG)

O Electroencefalograma (EEG) é uma técnica de exame cerebral usada desde 1929, quando o psiquiatra alemão Hans Berger descobriu que o cérebro gera uma actividade eléctrica capaz de ser registada [R10].

O EEG é realizado através da colocação de eléctrodos no escalpe de um indivíduo (como referenciado na fig.6). Os eléctrodos estão conectados a um poderoso amplificador de corrente eléctrica que permite aumentar a amplitude do sinal eléctrico, gerado pelo cérebro, milhares de vezes.



Fig.6 → Electroencefalograma (EEG)

O EEG, hoje em dia, tem utilização mais comum nas áreas de neurologia e psiquiatria, principalmente como auxílio no diagnóstico de doenças do cérebro, tais como epilepsia, distúrbios do sono e alguns tipos de tumores cerebrais.

O registo das ondas cerebrais através do EEG mostra que, de acordo com as tarefas realizadas, a actividade cerebral varia. O sinal EEG é caracterizado por ondas características que se destacam da actividade de fundo. Estas ondas são agrupadas por bandas de frequências.

Existem 5 tipos reconhecidos de actividade sinusoidal rítmica (bandas de frequências) no sinal EEG embora os limites para cada uma variem de artigo para artigo [R11].

- **Frequências Delta** - Estão associadas com o sono profundo. Algumas frequências na banda Delta libertam a hormona do crescimento humano (HGH), que intervém benéficamente na regeneração celular. A banda Delta está compreendida entre os 0.1 e 4Hz.
- **Frequências Teta** - Frequências compreendidas entre os 4 e 8Hz. Este estado cerebral propicia flashes de imagens do inconsciente e permite o acesso a memórias há muito tempo esquecidas.

As ondas Teta têm um papel importante em programas de modificação de comportamento e têm vindo a ser aplicadas no tratamento de vícios como o álcool e as drogas.

- **Frequências Alfa** - Frequências compreendidas entre os 8 e 13Hz. Estes tipos de frequências estão associadas a uma postura de relaxamento e meditação, sendo mais facilmente detectadas quando o sujeito se encontra de olhos fechados.

Dentro das frequências alfa existe uma variante de frequências a que se denominaram ritmos mu (10-12Hz). Estes ritmos encontram-se directamente relacionados com a intenção de movimento por parte do sujeito.

- **Frequências Beta** - Estas frequências encontram-se compreendidas entre os 13 e 30Hz e associam-se a um estado de concentração e atenção por parte de um indivíduo.

No estado Beta, os neurónios transmitem muito rapidamente as informações permitindo atingir assim estados de concentração elevados. O treino deste tipo de ondas é usado por terapeutas de biofeedback para tratamento de problemas de aprendizagem.

- **Frequências Gama** – As ondas Gama são ondas velozes e de alta frequência (30 a 50Hz), e são responsáveis por momentos de elevada criatividade, imaginação e concentração. Estão, por isso, associadas a um estado de actividade mental elevado.

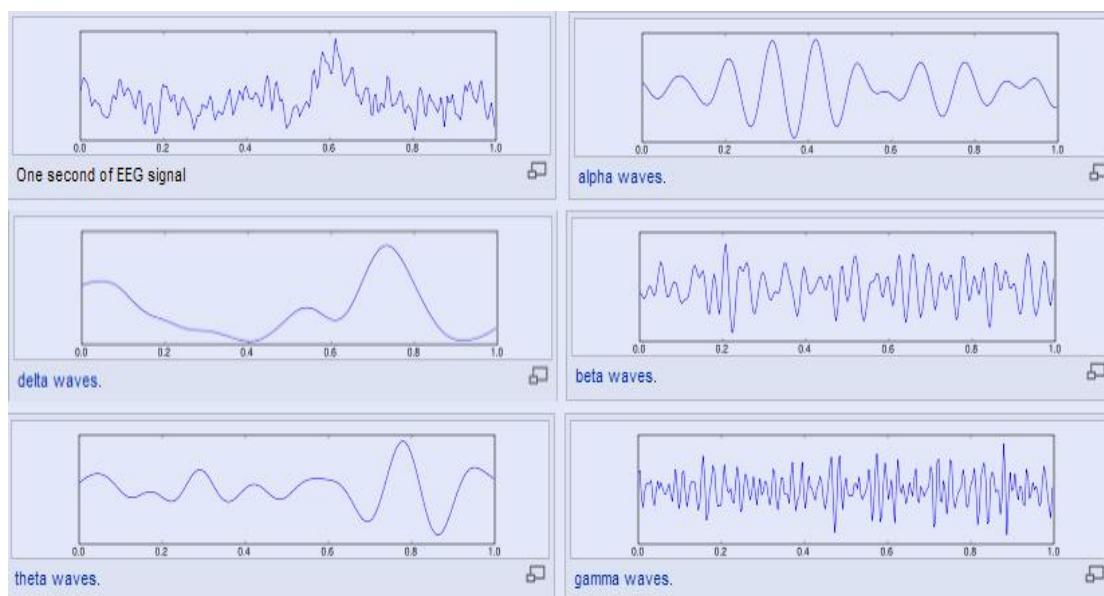


Fig.7 → Bandas de Frequência no Sinal EEG

2.5 Potenciais Evocados

Os Potenciais Evocados (PEs) podem ser definidos como respostas electrofisiológicas do córtex cerebral a estímulos sensoriais externos e têm sido usados, ao longo de décadas, para investigar doenças neurológicas.

Esta técnica, concebida por Dawson (1954), envolve a soma e a média das respostas electroencefalográficas a um estímulo sensitivo simples e repetitivo, que pode ser visual auditivo ou mesmo somatossentivo [R12]. Esta técnica tem como objectivo cancelar todos os outros tipos de actividade que ocorrem simultaneamente ao potencial evocado.

Os PEs alvo deste projecto, são os que resultam da reacção cerebral de um indivíduo durante a sua estimulação visual através de imagens, denominando-se por isso Potenciais Evocados Visuais (PEVs).

2.5.1 Potenciais Evocados Visuais (PEVs)

Os PEVs são predominantemente detectados na zona occipital através da estimulação visual do indivíduo.

É também de realçar que os PEVs deste projecto foram resultado do trabalho realizado na área de psicologia, denominado por “Differential effects of object-based attention on evoked potentials to fearful and disgusted faces”.

Nota: Enquanto, os potenciais evocados resultam de estímulos físicos ou fisiológicos, os potenciais evocados relativos a eventos (PERs), resultam de outro tipo de fenómenos, tais como, a memória, a expectativa, a atenção ou mesmo alterações no estado mental do indivíduo.

2.6 Classificadores

2.6.1 K Nearest Neighbors (KNN)

Em reconhecimento padrão o algoritmo KNN é um método de classificação de objectos, que tem por base os exemplos de treino mais próximos em função do espaço existente entre estes e o objecto a classificar.

Neste algoritmo um objecto é classificado por uma maioria de votos dos seus vizinhos, com o objecto a ser atribuído à classe mais comum entre os seus K vizinhos mais próximos. K é, por norma, um número inteiro positivo ímpar, de forma a evitar empates e, por consequente, indefinições na classificação do objecto. Para as situações em que K=1, o objecto a classificar é simplesmente atribuído à classe do seu vizinho mais próximo.

Os vizinhos são retirados de um conjunto de objectos para os quais a classificação é já conhecida, o que pode ser visto como o conjunto de treino do algoritmo. Para identificar os seus vizinhos os objectos são representados através dos seus vectores posição no espaço multidimensional. No âmbito de se calcularem e compararem as distâncias entre o objecto a classificar e os seus vizinhos (já devidamente classificados), foi utilizada a distância Euclidiana sendo também válida a utilização de outras distâncias como a distância de Manhattan [R13].

A distância Euclidiana entre dois pontos X e Y é dada por:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}.$$

Equação 1 -> Distância Euclidiana entre dois pontos

Como no decorrer do projecto foi utilizada a ferramenta **stprtool**, que faz automaticamente os cálculos das distâncias entre o objecto a classificar e os seus vizinhos, fica aqui um exemplo do funcionamento do classificador KNN [14].

Exemplo:

O objectivo deste exemplo é classificar um tipo de pilhas (Z), como boas ou más, com base no seu preço e na sua durabilidade, considerando outros 4 tipos de pilhas (A, B, C, D). As notas tanto para a durabilidade como para o preço variam de 1 (bom) a 7 (mau).

Pilhas	X1 = Durabilidade	X2 = Preço	Y = Classificação
A	7	7	Má
B	7	4	Má
C	3	4	Boa
D	1	4	Boa

Tabela 1 -> Explicação do Algoritmo KNN (1)

Em relação ao tipo de pilhas a classificar sabe-se que este tipo de pilhas apresenta durabilidade = 3 e preço = 7. Nas tabelas seguintes, é feito o procedimento para averiguar, com estas características, se este tipo de pilhas tem boa ou má classificação. Neste exemplo são usados 3 vizinhos (K=3).

Pilhas	X1 = Durabilidade	X2 = Preço	Distância para o objecto em estudo	Ranking de afastamento	Está incluído nos 3 vizinhos?
A	7	7	4	3	Sim
B	7	4	5	4	Não
C	3	4	3	1	Sim
D	1	4	$\sqrt{13}$	2	Sim

Tabela 2 -> Explicação do Algoritmo KNN (2)

A partir deste procedimento de cálculo de distâncias, pode-se concluir que as pilhas do tipo Z a classificar têm uma boa classificação visto que têm mais vizinhos com boa classificação (2 vizinhos) do que com má classificação (1 vizinho).

Para além deste classificador a ferramenta **stprtool** apresenta também outros classificadores como é o caso do SVM (Support Vector Machines), classificador que irá ser utilizado na tese referente à Identificação Biométrica.

É ainda de referir que este algoritmo encontra-se entre os mais simples algoritmos de aprendizagem.

Capitulo 3

Base de Dados e Métodos

3.1 Base de Dados

3.1.1 Resumo da Experiência

Este projecto tem como estrutura, a base de dados utilizada num trabalho no âmbito da neuropsicologia referenciado em [R15].

Nesta base de dados foram gravados os sinais EEG de 16 pessoas quando expostas a estímulos visuais.

Os estímulos visuais são compostos por 48 imagens de faces e 16 imagens de casas. As faces provêm de 16 indivíduos diferentes (8 homens e 8 mulheres), onde cada um apresenta 3 expressões faciais diferentes (normal, medo e nojo), sendo seleccionadas da base de dados de Ekman and Friesen (Ekman & Friesen, 1976) e de uma base de dados similar, criada com os mesmos critérios, na Universidade Autónoma de Madrid. As casas foram seleccionadas de vários sites de internet.

Neste trabalho, cada imagem facial foi sobreposta com uma casa sendo depois formados 4 blocos experimentais. Em dois desses blocos foi pedido aos participantes que se concentrassem nas faces e nos outros dois foi-lhes pedido que se concentrassem nas casas. Cada um destes blocos era composto por 256 imagens (8 indivíduos x 8 casas x 2 expressões mostradas 2 vezes cada). A tarefa de cada participante era determinar, em cada imagem, se o estímulo relevante (cara ou casa dependendo do bloco em questão) era igual ao apresentado na imagem anterior, não sendo relevante a expressão facial.

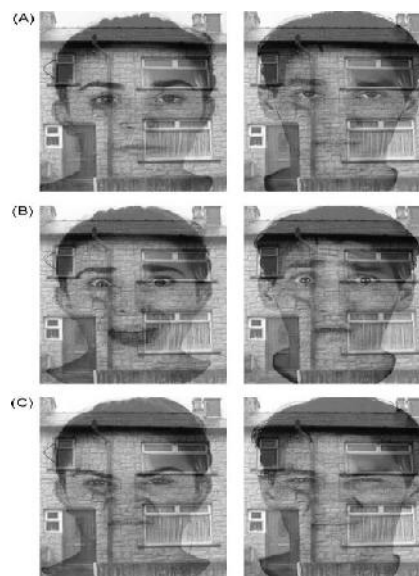


Fig.8 -> Exemplo de Imagens utilizadas

Cada imagem foi apresentada aos participantes durante 300ms, sendo imediatamente precedida por um “fixation cross”, que era visualizado pelos participantes durante os seguintes 500ms. O intervalo entre imagens foi de 2000ms. O tempo considerado para cada imagem teve por base os estudos existentes no ramo.

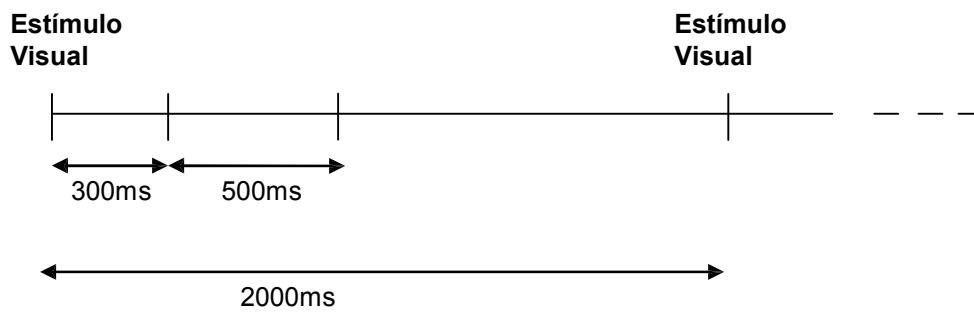


Fig.9-> Período de duração dos Estímulos

3.1.2 Extracção dos Resultados da Experiência

Os sinais EEG, referentes às 16 pessoas que se submeteram ao estudo, foram adquiridos usando um conjunto de 20 eléctrodos (Fp1-2, F3-4, C3-4, P3-4, O1-2, F7-8, T3-4, T5-6, Fz, Cz, Pz e Oz), e foram gravados a uma taxa de 250 amostras por segundo.

Foi também gravado o sinal EOG através dos canais oculares (Cb1 e Cb2). Este sinal é utilizado como uma espécie de referência visto que é comparado com o sinal EEG, resultando desta comparação, uma primeira remoção de artefactos do sinal EEG.

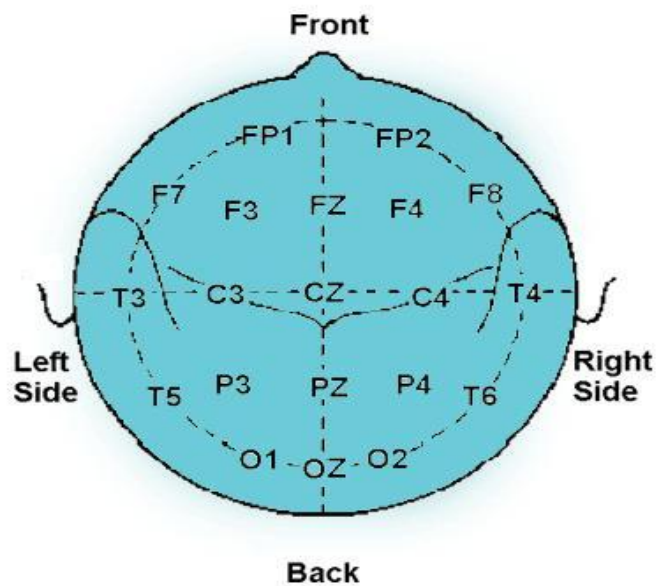


Fig.10 -> Localização dos Canais

É também de referir que, ainda no processo que envolve a obtenção dos dados a serem trabalhados neste projecto, foram utilizados os seguintes elementos:

- Impedância abaixo dos 5K Ω .
- Filtro passa banda entre os 0.05 e 30 Hz.
- “Notch Filter” nos 50 Hz de forma a tirar a possível interferência da rede.
- Um conversor A/D de 12-bit.

Para cada imagem apresentada, apenas 1600ms de sinal EEG ficaram registados. Destes 1600ms, 148ms (37 amostras do sinal) correspondem a pré estímulo seguindo-se os restantes 1452ms pós estímulo. Esta porção de sinal irá ser tratada por Época ou Evento no decorrer deste projecto.

Assim, os dados trabalhados neste projecto, são compostos por segmentos sincronizados com estímulo, ou seja, o sinal trabalhado neste projecto trata-se de um sinal descontinuado no tempo.

3.2 Processamento do sinal EEG

Depois de adquiridos os dados através da experiência previamente mencionada, foi necessário processar o sinal EEG de forma a extrair a informação relevante para o processo de autenticação.

A figura seguinte (fig.11) ilustra os passos utilizados para o processamento do sinal.



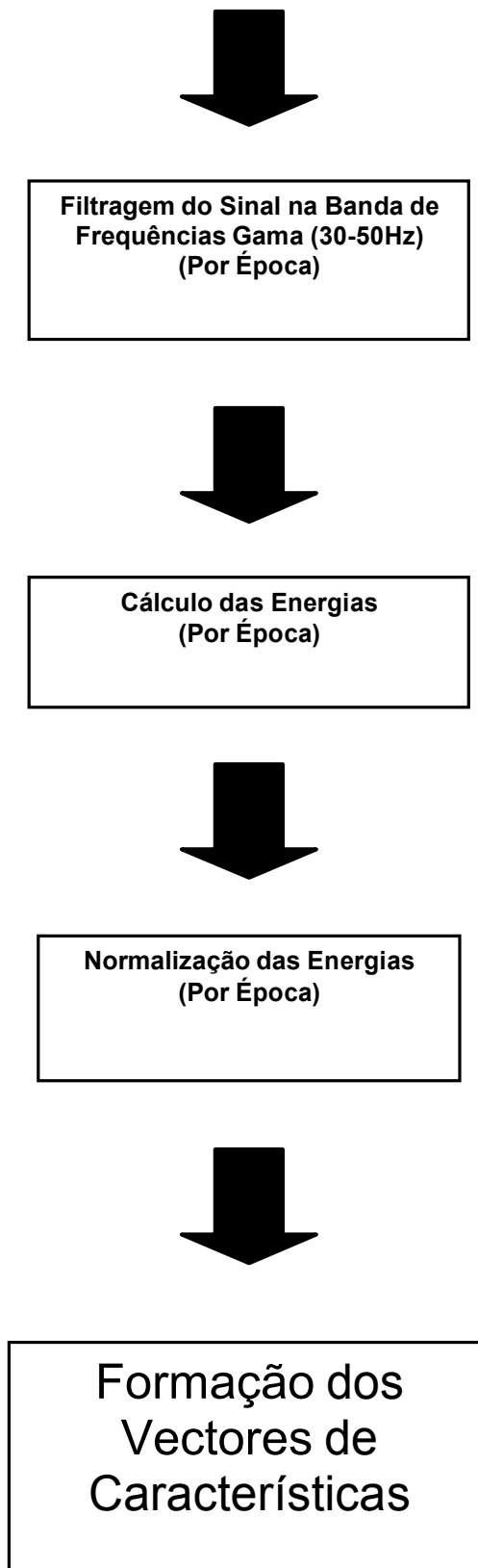


Fig.11 -> Fases do Processamento do Sinal EEG

3.2.1 Remoção das Épocas com Artefactos

O EEG (Electroencefalograma), como já referido, é uma técnica médica onde é gravada a actividade cerebral de um indivíduo.

Embora seja concebido para este propósito (gravar actividade cerebral), o EEG grava também actividades eléctricas resultantes de outros pontos que não o cérebro.

Assim, Artefactos em Electroencefalografia, são definidos como todo o potencial eléctrico proveniente de outra fonte que não o cérebro, mas que é captado pelos eléctrodos, constituindo assim um dos principais problemas na análise do EEG [R16].

Os artefactos podem ser divididos em duas categorias, fisiológicos e não fisiológicos. Os fisiológicos são gerados pela própria pessoa sujeita ao EEG, como é o caso do piscar de olhos, movimentos oculares, musculares, respiratórios e batimento cardíaco. Os artefactos não fisiológicos, como a má conectividade dos eléctrodos ou mesmo a proximidade de dispositivos eléctricos, não estão directamente relacionados com a pessoa em análise.

Neste projecto, e já depois de serem removidos alguns artefactos com base no sinal EOG, os restantes artefactos foram removidos a olho, sempre que era observado um pico de tensão ou anomalia.

A fig.12, capturada a partir do **EEGLAB**, ilustra o potencial eléctrico relativo à **pessoa 1** quando esta foi sujeita a estímulos visuais faciais. Nesta figura é possível observar o potencial eléctrico (em mv), da primeira à quinta época, ao longo dos 22 (20+2) canais.

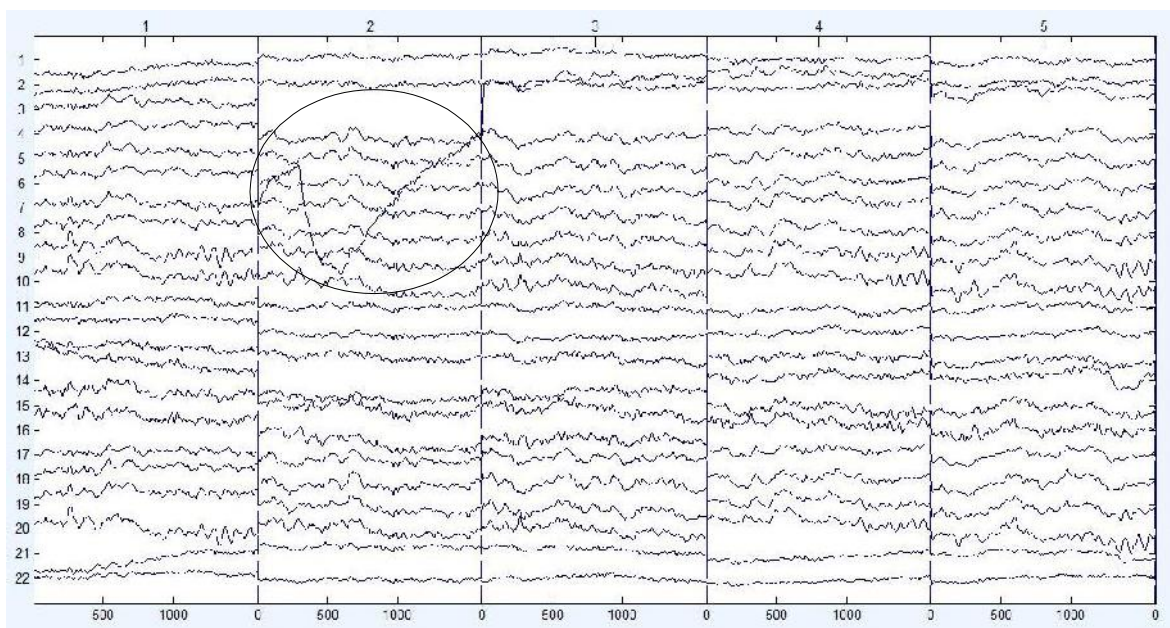


Fig.12 -> Potencial Eléctrico ao longo de 5 Épocas

Como na segunda época se verificou a presença de um artefacto, essa época foi removida tal como se pode comprovar com o auxílio da fig.13.

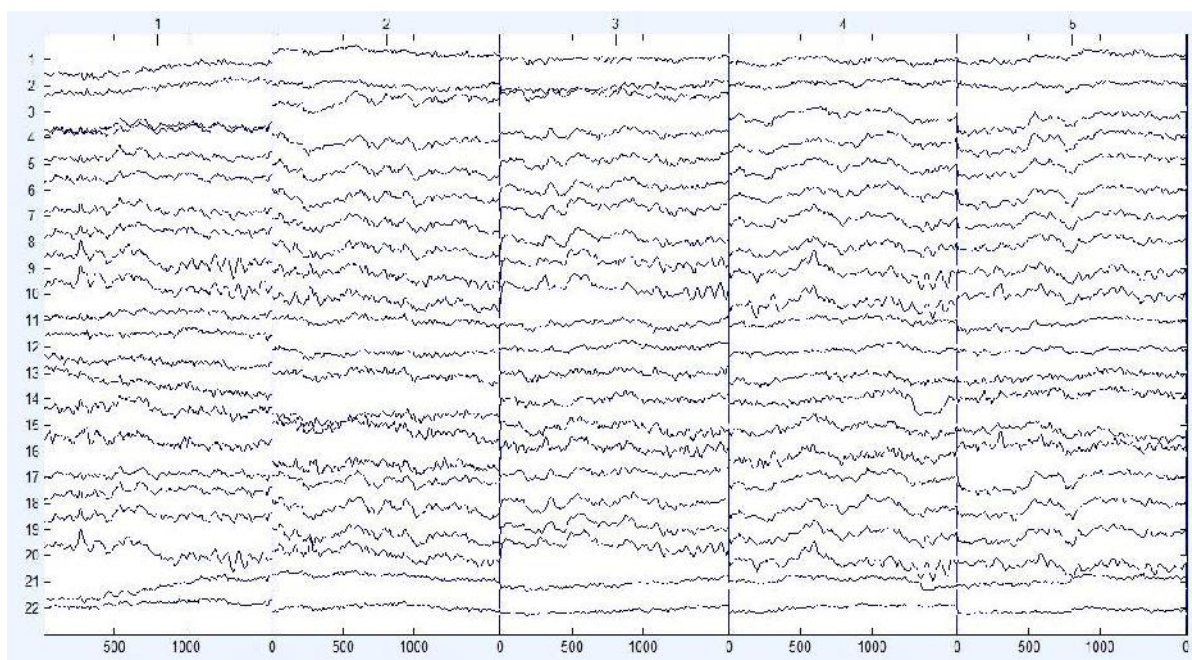


Fig.13 -> Potencial Eléctrico ao longo de 5 Épocas (já com a segunda época removida)

Para cada uma das pessoas utilizou-se uma função (**fwplread**) para verificar o índice correspondente às épocas com artefactos. Conhecido o número e a localização dos artefactos procedeu-se à sua remoção através do uso de outra função (**rejeicao**).

3.2.2 Filtragem do Sinal na Banda de Frequências Gama

Nesta fase, dos sinais EEG já sem artefactos, é extraída uma parte da banda de frequências gama (30-50) Hz (função **Filtro**). Este processo é realizado **época a época**.

Escolheu-se esta banda de frequências visto estar associada, como já anteriormente referido, à criatividade, imaginação e concentração de um indivíduo, ou seja, a um estado de actividade mental elevado.

Para a extracção desta banda de frequências recorreu-se, neste projecto, a um filtro butterworth passa-banda de ordem 4 que apresenta um decaimento de -80 dB por década. A escolha deste filtro justifica-se, visto que este apresenta uma resposta em frequência muito plana (não possui ripple na banda passante) e aproxima-se do zero na banda rejeitada. Para além disso, é o único filtro que mantém o mesmo formato para ordens mais elevadas.

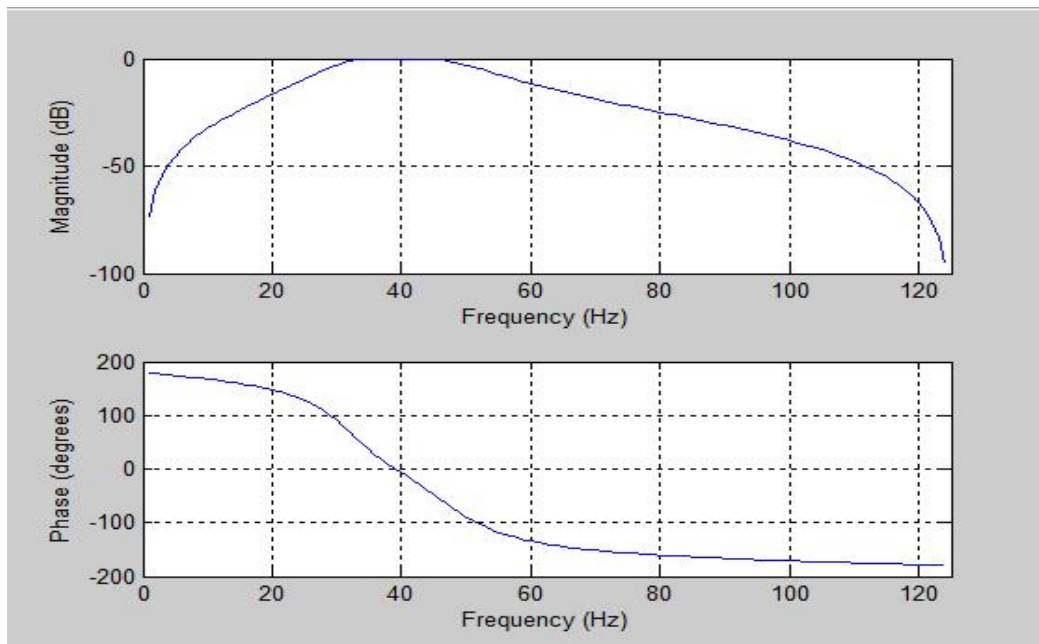


Fig.14 -> Resposta em Magnitude e Fase do Filtro Butterworth de Ordem 4

Ainda de referir:

- Neste projecto foram testados filtros butterworth de ordem 6 e de ordem 8, em detrimento do filtro de ordem 4, na banda de frequências gama.
- Foi também testada a utilização do filtro butterworth na banda de frequências beta, visto que estas frequências estão também associadas a estados de concentração elevados.

3.2.3 Cálculo e Normalização da Energia para cada Época

Nesta fase é calculada a energia do sinal EEG (utilizando a função **energia**), já filtrado, para cada canal (20 canais). Este cálculo, **realizado por época**, é efectuado, tendo por base o teorema da equivalência tempo-frequência de Parseval.

$$\frac{1}{N} \sum_{i=0}^{N-1} |\text{FFT amplitude}[i]|^2 = \sum_{i=0}^{N-1} |\text{data magnitude}[i]|^2$$

Equação 2 -> Cálculo das Energias

A tabela 3 representa um exemplo do cálculo das energias para a **Pessoa 1**. Estas energias, calculadas para cada canal, estão associadas à primeira época dos estímulos referentes às faces.

Canal	Energia	Canal	Energia	Canal	Energia	Canal	Energia
FP1	246.2739	C4	222,3807	F7	294,4567	T6	706,4517
FP2	196.2101	P3	278,9640	F8	269,3052	Fz	214,3728
F3	286.0647	P4	279,8255	T3	580,5088	Cz	237,8330
F4	246,4250	O1	377,5455	T4	803,4865	Pz	243,3509
C3	259,8652	O2	336,5947	T5	685,0329	Oz	298,8147

Tabela 3 -> Valores arredondados das Energias por Canal (primeira Época)

Finalmente, e já com as energias calculadas, procede-se à normalização das mesmas. Nesta fase, para cada época, é dividida a energia correspondente a cada canal pela energia total da época em questão.

Na tabela 4 são normalizadas as energias apresentadas na tabela anterior.

Canal	Energia	Canal	Energia	Canal	Energia	Canal	Energia
FP1	0,0349	C4	0,0315	F7	0,0417	T6	0,1000
FP2	0,0278	P3	0,0395	F8	0,0381	Fz	0,0303
F3	0,0405	P4	0,0396	T3	0,0822	Cz	0,0337
F4	0,0349	O1	0,0534	T4	0,1137	Pz	0,0345
C3	0,0368	O2	0,0477	T5	0,0970	Oz	0,0423

Tabela 4 -> Valores arredondados das Energias Normalizadas por Canal (primeira Época)

3.2.4 Formação e Análise dos Vectors das Características

Nesta fase é possível observar o resultado do processo atrás descrito, ou seja, a formação de um conjunto de vectores de características para cada pessoa. O número

destes vectores varia de pessoa para pessoa visto que, devido à remoção de épocas com artefactos, o número de épocas varia também de pessoa para pessoa.

A fg.15 ilustra o vector de características, referente à primeira época, para a **Pessoa 1**.

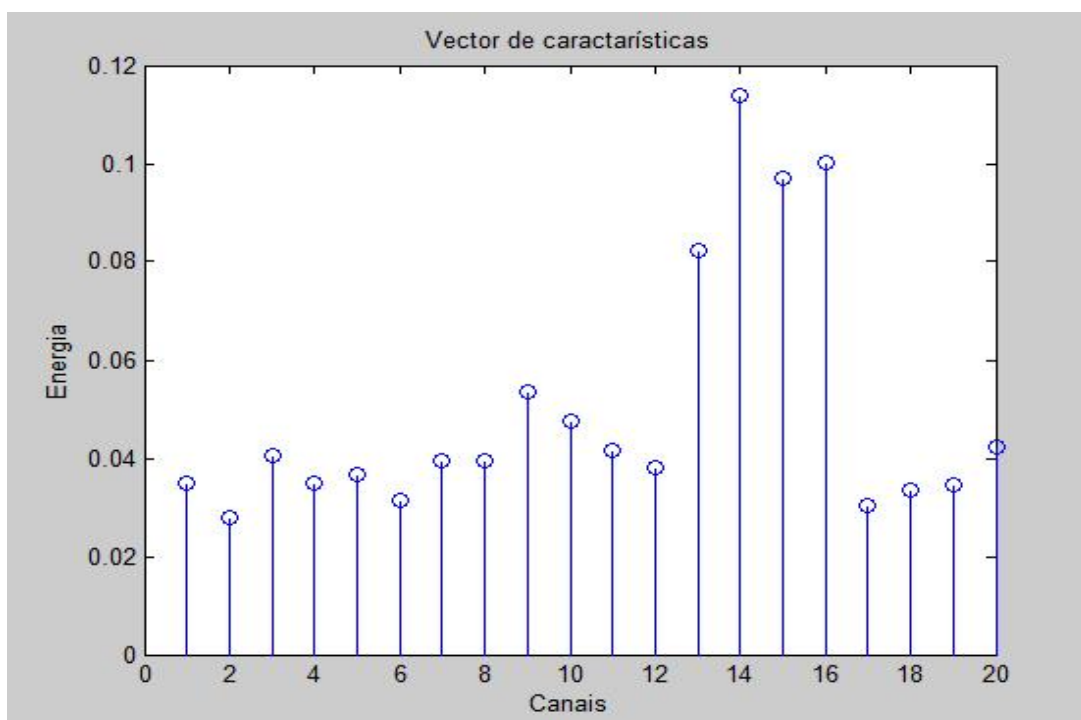


Fig.15 -> Vector de Características da Pessoa 1 (referente à primeira Época)

Com os vectores de características já determinados, o objectivo nesta fase é verificar se estes apresentam o mesmo padrão para cada indivíduo e padrões diferentes entre indivíduos diferentes.

É ainda de referir que das 16 pessoas apenas se utilizaram os dados referentes a 13 destas pessoas, devido ao número de eventos rejeitados de 3 destas pessoas ser demasiado elevado.

Nas figuras seguintes (fig.16, fig.17 e fig.18) são apresentados os vectores de características referentes às **Pessoas (1, 2 e 3)**. Estes vectores de características encontram-se relacionados apenas com os estímulos referentes às faces.

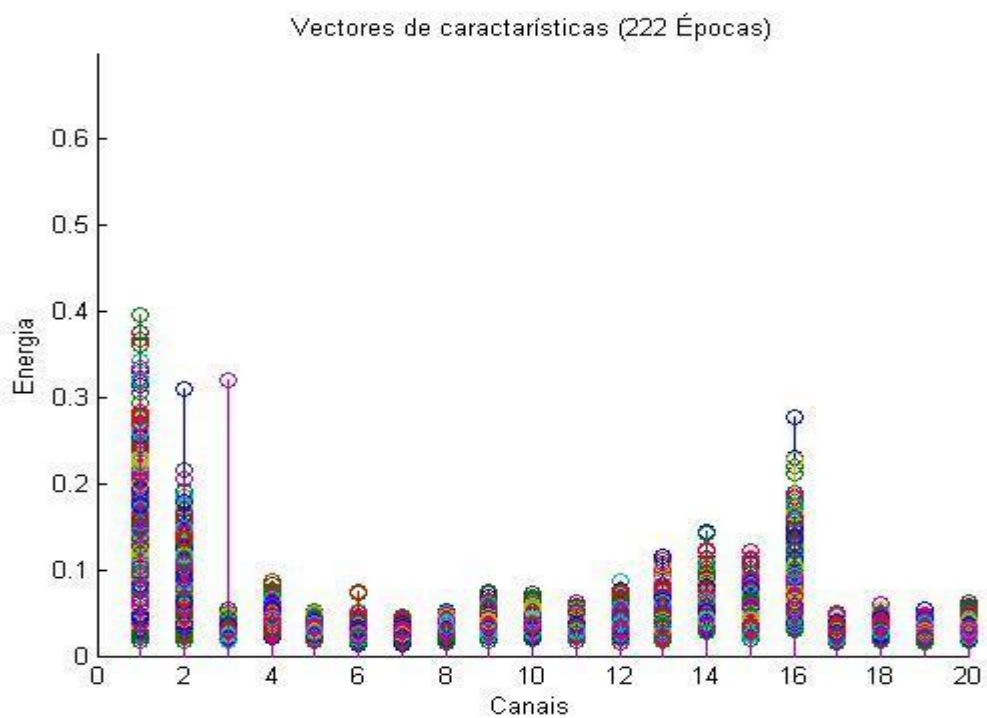


Fig.16 -> Vectores de Características referentes à Pessoa 1

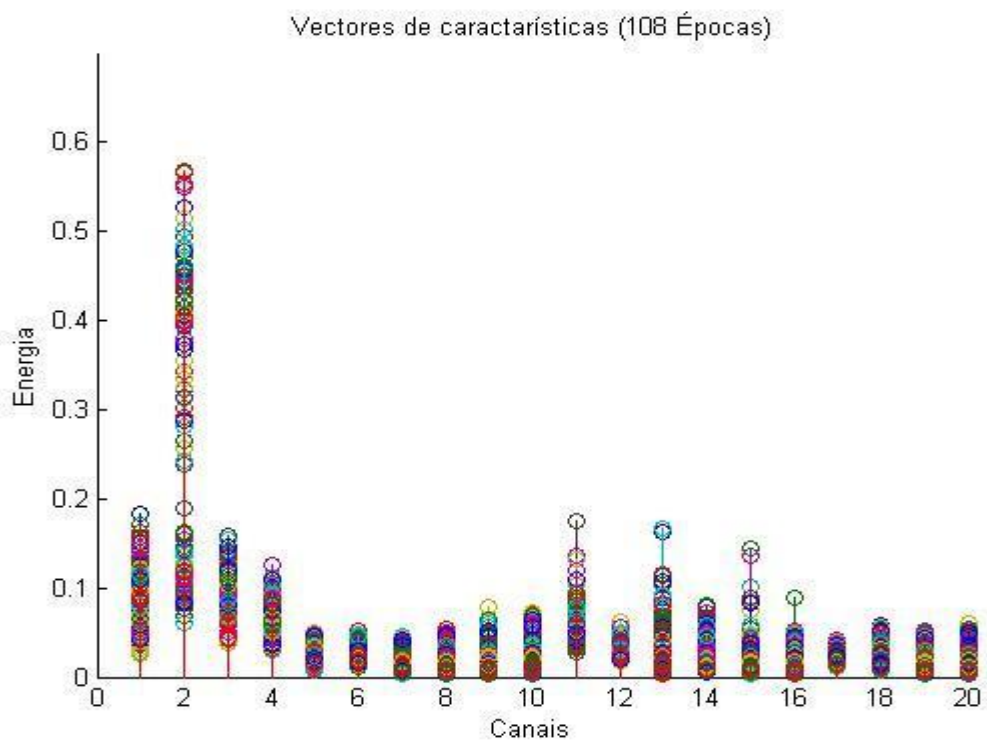


Fig.17 -> Vectores de Características referentes à Pessoa 2

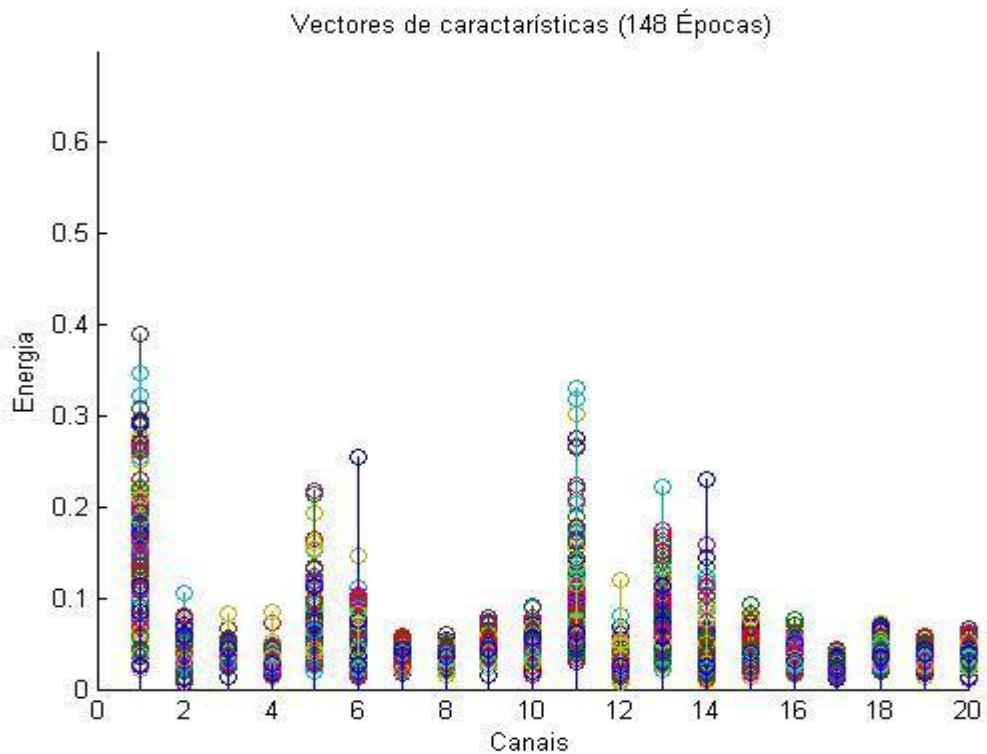


Fig.18 -> Vectores de Características referentes à Pessoa 3

Analisando as figuras anteriores (16,17 e 18), verifica-se que, após a apresentação de vários estímulos visuais, os vetores de características formados apresentam o mesmo padrão para o mesmo indivíduo e padrões diferentes para indivíduos diferentes. Ao terminar o processamento de sinal esta seria a conclusão que se esperaria obter.

Capitulo 4

Classificação e Estudos Efectuados

4.1 Classificação

Ao longo do capítulo anterior, foi apresentado todo o processo que envolve o processamento do sinal EEG e que termina com a extracção dos vectores de características referentes aos 13 indivíduos em estudo.

Como já anteriormente referido, o número destes vectores, varia de indivíduo para indivíduo. Isto acontece devido ao número de épocas (com artefactos) removidas, ser diferente consoante o indivíduo em questão.

Cada um destes vectores de características corresponde a uma determinada época e, é composto por valores de energia normalizada do sinal EEG, na banda gama (30-50) Hz, para cada um dos 20 canais utilizados.

Em seguida, será aplicado o classificador KNN aos vectores de características extraídas para cada indivíduo, no âmbito de verificar a viabilidade destas características e do classificador numa possível posterior utilização num sistema de autenticação biométrico.

4.1.2 Casos de Estudo

Neste projecto foram utilizados dois cenários diferentes para se realizar a classificação dos vectores de características, utilizando o classificador KNN.

1. Foram utilizadas, para as 13 pessoas, as épocas relativas aos estímulos faciais para o treino do classificador, sendo as épocas dos estímulos relativos às casas utilizadas para testar o classificador (função -> **func_classknn**).
2. O classificador foi treinado, para as 13 pessoas, com metade das épocas relativas aos estímulos das casas e metade das épocas relativas aos estímulos faciais. O teste do classificador foi realizado recorrendo-se à outra metade dos estímulos (casas e faces) também para as 13 pessoas (função -> **func_classknn_mix**).

Neste ponto, é também importante salientar outras mudanças operadas, antes e durante a classificação, tendo em vista verificar como variam os resultados.

- Alteração da banda de frequências utilizada
- Alteração da ordem do filtro utilizado

- Redução progressiva do número de épocas utilizadas para treino e teste do classificador
- Redução progressiva do número de canais utilizados
- Alteração do valor de K do classificador

4.2 Pós Processamento (P.P.) da saída do Classificador

Antes de se falar propriamente nos resultados obtidos, é importante falar no que foi feito para os melhorar. Assim, no âmbito de diminuir o erro do classificador, criou-se uma função (**pos_proc_alterado**) que tem como único propósito fazer um Pós Processamento da sequência temporal de valores à saída do classificador, isto é, incrementar progressivas alterações na sequência temporal de valores que resulta da classificação.

Estas alterações são então incrementadas progressivamente, sob a forma de iterações, e visam reduzir o erro de autenticação final, ou seja, corrigir o classificador.

Passa-se então a explicar o funcionamento e procedimento do Pós Processamento dos resultados.

A tabela seguinte (tabela 5) representa uma parte da sequência temporal de valores à saída do classificador (ypred). Esta parcela da sequência temporal de valores diz respeito à pessoa3, e sendo assim, sempre que se observa um número diferente de 3, é sinal que o classificador se enganou a classificar a pessoa.

ypred <32042 double>														
392	393	394	395	396	397	398	399	400	401	402	403	404	405	406
3	4	13	3	3	3	3	3	5	7	11	3	4	3	3


Tabela 5 -> Parte da sequência temporal de valores à saída do classificador (ypred)

Assim, e de forma a procurar corrigir o classificador, a função criada vai, numa **primeira iteração**, correr esta parcela de valor em valor analisando o valor em questão e os dois seguintes, como se de uma janela flutuante se tratasse. Para este exemplo, esta janela flutuante vai analisar em primeira instância os valores com os índices 392, 393 e

394, depois os valores com os índices 393, 394 e 395, depois os valores com os índices 394, 395 e 396 e assim sucessivamente.

Por cada 3 valores, o primeiro é comparado com o terceiro. Se estes dois forem iguais e, se o valor do meio for diferente, este último é alterado, passando a ser igual ao primeiro e terceiro valores.

Na tabela seguinte, é possível observar a primeira iteração efectuada para (ypred), dando assim origem a uma nova sequência temporal de valores (ypred1).



ypred1: <1x2042 double>														
392	393	394	395	396	397	398	399	400	401	402	403	404	405	406
3	4	3	3	3	3	3	3	7	11	3	3	3	3	3

Tabela 6 -> Pós Processamento (primeira iteração)


Comparando ypred e ypred1, verifica-se que foi efectuada a correcção de um valor (índice 403).

É ainda de realçar que, este procedimento funciona como se tivesse memória, ou seja, se um valor for alterado, esta alteração vai-se manter quando a janela flutuante passar para a próxima análise. Por exemplo, quando são analisados os valores com os índices 402, 403 e 404, é alterado o valor 403 para 3 e, em seguida, quando são analisados os valores 403, 404 e 405, o valor com o índice 403 continuará a apresentar o valor 3.

Assim o Pós Processamento prossegue com a **segunda iteração**. O processo é idêntico ao da primeira iteração só que, neste caso, a sequência temporal de valores (ypred1) irá ser percorrida valor a valor, sendo analisado o valor em questão e os 3 seguintes.

Nesta situação o primeiro valor é comparado com o quarto. Se estes dois forem iguais entre si e, se o segundo e o terceiro forem diferentes dos dois primeiros, estes dois últimos vão ser alterados e vão ficar com o mesmo valor dos primeiros (primeiro e quarto valores).

Na tabela seguinte está representada parte da nova sequência temporal de valores (ypred2).




ypred2: <1x2042 double>														
392	393	394	395	396	397	398	399	400	401	402	403	404	405	406
3	3	3	3	3	3	3	3	7	11	3	3	3	3	3

Tabela 7 -> Pós Processamento (segunda iteração)

Comparando ypred1 e ypred2, verifica-se que mais dois valores (393 e 394) foram corrigidos.

Partindo agora de ypred2 e, continuando o Pós Processamento com uma **terceira iteração**, obtém-se uma nova sequência temporal de valores (ypred3).



ypred3: <1x2042 doub e>														
392	393	394	395	396	397	398	399	400	401	402	403	404	405	406
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Tabela 8 -> Pós Processamento (terceira iteração)

O processo utilizado na execução desta nova iteração é semelhante ao já utilizado nas outras duas, com a diferença que neste caso os valores são analisados 5 a 5. Mais uma vez, se o primeiro e neste caso o quinto valores forem iguais e, os outros 3 do meio forem diferentes destes dois, então estes últimos 3 (segundo, terceiro e quarto valores) passarão a ter o mesmo valor dos dois primeiros (primeiro e quinto valores).

Comparando desta vez ypred2 e ypred3, mais uma vez se verifica uma correcção do classificador, visto que são corrigidos mais 3 valores (399, 400 e 401).

Assim, neste caso, desde a primeira à terceira iterações, foram corrigidos 6 valores, tornando-se então possível reduzir o erro de autenticação, ou seja, obter uma melhor classificação por parte do classificador.

Neste exemplo não é possível observar mas, é importante referir que, neste projecto, recorreu-se sempre a 5 iterações na parte de Pós Processamento sendo que, na **quarta iteração**, torna-se possível a correcção de 4 valores e, na **quinta iteração**, de 5 valores. Assim, o gráfico seguinte (fig.19) ilustra a correcção do erro, com as 5 iterações (Pós Processamento completo), quando é percorrida a sequência temporal de valores completa à saída do classificador.

Evolução do Erro ao longo do P.P.

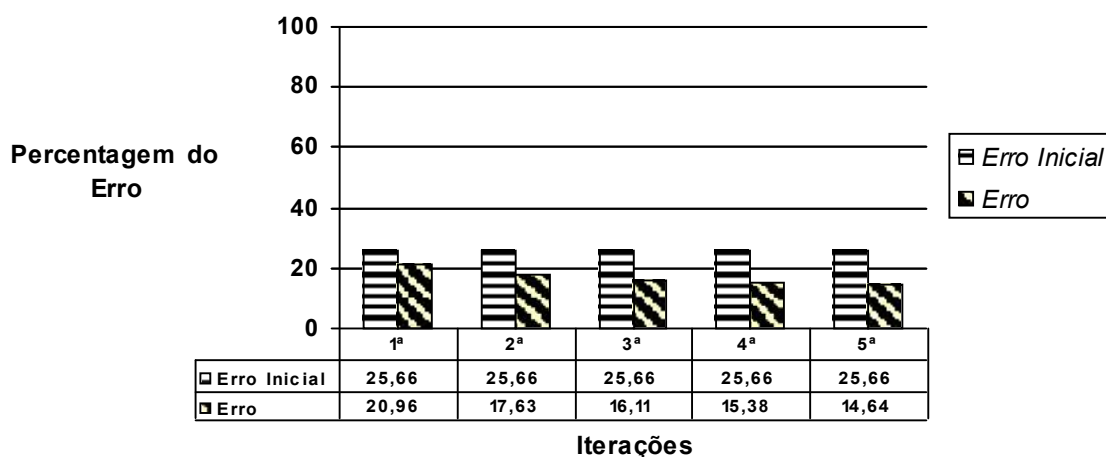


Fig.19 -> Evolução do Erro ao longo do P.P. completo

Sobre a fig.19 é de referir que o erro inicial corresponde ao erro do classificador sem P.P. Este erro aparece em todas as iterações apenas como elemento comparativo.

Porém, o Pós Processamento pode induzir em erro o classificador.

As tabelas seguintes representam uma parte da sequência temporal de valores (ypred), resultante do teste do classificador, e uma parte da sequência temporal de valores (ypred5), que resulta do Pós Processamento completo de ypred. Esta sequência temporal de valores diz respeito à pessoa 4, logo, sempre que se avistar um valor que defira de 4 é sinal de má classificação por parte do classificador.

ypred <32042 double>														
608	609	610	611	612	613	614	615	616	617	618	619	620	621	622
4	4	4	4	4	4	4	1	1	1	4	4	1	1	4

Tabela 9 -> Parte da sequência temporal de valores à saída do classificador (ypred)

ypred5 <32042 double>														
608	609	610	611	612	613	614	615	616	617	618	619	620	621	622
4	4	4	4	4	4	4	1	1	1	4	4	1	1	4

Tabela 10 -> Parte da sequência temporal de valores resultante do Pós Processamento completo (ypred5)

Como é possível observar, através das tabelas anteriores (tabelas 9 e 10), nesta situação, o Pós Processamento dos resultados é prejudicial, isto porque apenas um valor que padecia de má classificação foi correctamente corrigido (610), enquanto que dois valores que padeciam de classificação correcta foram induzidos em erro (618 e 619).

Pese embora o facto desta fase (Pós Processamento), em alguns casos, poder induzir em erro o classificador, na maioria das situações torna-se benéfico e permite reduzir significativamente o erro da classificação. É também importante referir que em algumas situações o erro poderá aumentar de uma iteração para outra, não sendo contudo muito comum.

A importância do Pós Processamento irá ser evidenciada mais à frente na parte que diz respeito ao erro de classificação individual, pois irá ser possível observar a redução do erro para cada pessoa. Na apresentação dos resultados, estes vão ser apresentados com e sem Pós Processamento.

4.3 Estudos efectuados

Como já anteriormente referido, foram utilizados inicialmente dois cenários distintos no âmbito proceder ao treino e ao teste do classificador. Foram também efectuadas experiências com o propósito de otimizar os resultados obtidos constando, entre as quais, a alteração da ordem do filtro, a alteração da banda de frequências, a redução progressiva do número de canais e do número de épocas e a alteração do valor de K do classificador.

Em seguida, vão ser apresentados os resultados obtidos no decorrer deste projecto, ou seja, a percentagem de erro associada ao classificador KNN, após classificar os vectores de características usados para simulação do classificador.

Em primeiro lugar, foi importante verificar, qual dos cenários utilizados para o treino e teste do classificador, traduz melhores resultados.

4.3.1 Treino e Validação com Tipos de Estímulos Distintos

Fazendo recurso das épocas relativas aos estímulos faciais para o treino do classificador e das épocas relativas aos estímulos das casas para o teste do mesmo (Cenário1) obtém-se:

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 36,08%
Percentagem de erro após Pós Processamento = 24,04%

Fig.20 -> Treino e Validação com tipos de Estímulos distintos

4.3.2 Treino e Validação com Tipos de Estímulos Semelhantes

Recorrendo a metade das épocas relativas aos estímulos das casas e metade das épocas relativas aos estímulos faciais para o treino e, à outra metade dos estímulos (casas e faces) para o teste do classificador (Cenário 2), obtém-se:

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 25,66%
Percentagem de erro após Pós Processamento = 14,64%

Fig.21 -> Treino e Validação com tipos de Estímulos semelhantes

Pela análise dos resultados anteriores, verifica-se que o segundo cenário é aquele que permite obter melhores resultados na classificação, visto que, é o que apresenta menor percentagem de erro.

Recorrendo sempre a este segundo cenário para treino e teste do classificador, procedeu-se a algumas experiências/testes (seguidamente apresentadas), com o objectivo de otimizar o processo de classificação.

4.3.3 Variação da Ordem do Filtro

- Testou-se um filtro Butterworth de ordem 6 para obter os vectores de características
- Testou-se um filtro Butterworth de ordem 8 para obter os vectores de características

Utilizando:

20 Canais activos
Filtro de ordem 6 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 27,72%
Percentagem de erro após Pós Processamento = 14,59%

Fig.22 -> Variação da Ordem do Filtro (Ordem 6)

Utilizando:

20 Canais activos
Filtro de ordem 8 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = **28,21%**

Percentagem de erro após Pós Processamento = **13,71%**

Fig.23 -> Variação da Ordem do Filtro (Ordem 8)

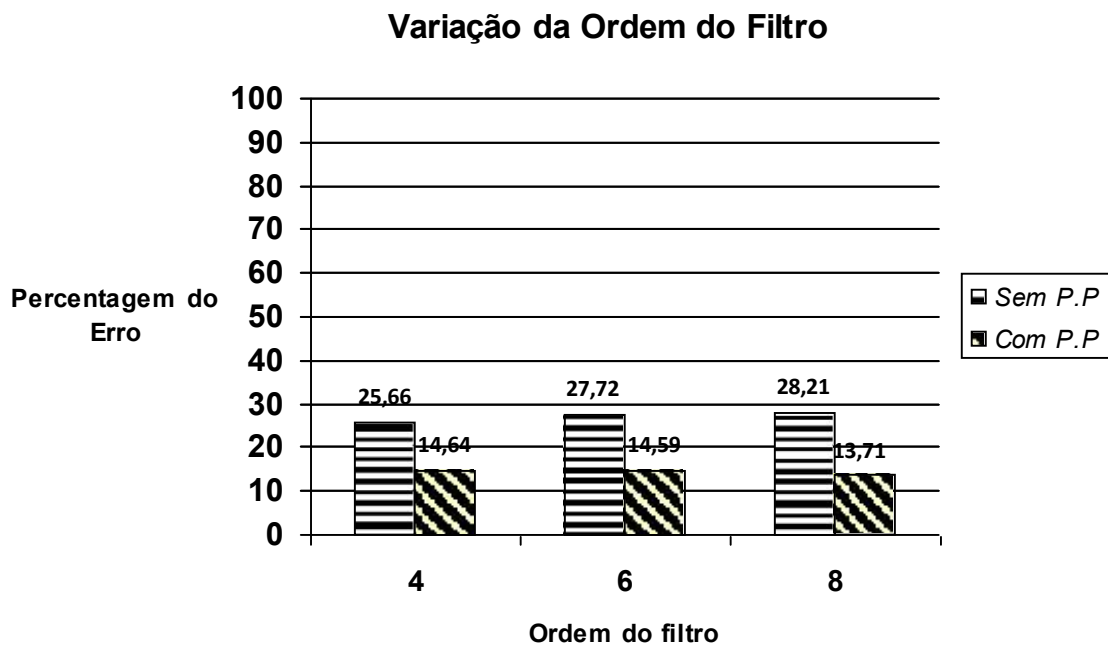


Fig.24 -> Variação da Ordem do Filtro

Pelos resultados obtidos (fig.22 e fig.23), constata-se que, sem Pós Processamento, o aumento gradual da ordem do filtro, vai provocar um ligeiro aumento, também gradual, do erro. Com o auxílio da fig.24 é possível fazer uma comparação gráfica do erro, com e sem P.P, para os filtros de diferentes ordens.

Apesar dos filtros de ordem 6 e 8 produzirem ligeiramente melhores resultados após Pós Processamento (neste exemplo), o filtro de ordem 4 produz melhores resultados no aspecto geral.

Para além disso, quanto menor for a ordem do filtro mais suave é a sua resposta e mais simples é de implementar. Por consequente, nas experiências/testes seguintes, foi utilizado o filtro de ordem 4.

4.3.4 Variação da Banda de Frequências Extraída

- Foi testada a banda de frequências beta

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências beta
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = **28,99%**

Percentagem de erro após Pós Processamento = **14,99%**

Fig.25 -> Resultados para a Banda de Frequências Beta

Utilizando a banda de frequências beta como banda de frequências que se extrai do sinal EEG e analisando os resultados daí retirados (fig.25) verifica-se um ligeiro aumento do erro de classificação. Esta pequena diferença de valores pode ser visualizada graficamente com o auxílio da figura seguinte (fig.26).

Esta ligeira diferença verifica-se também para os pontos que se seguem (redução do número de canais e épocas). Por esta razão, irá continuar a utilizar-se a banda de frequências gama nos pontos seguintes.

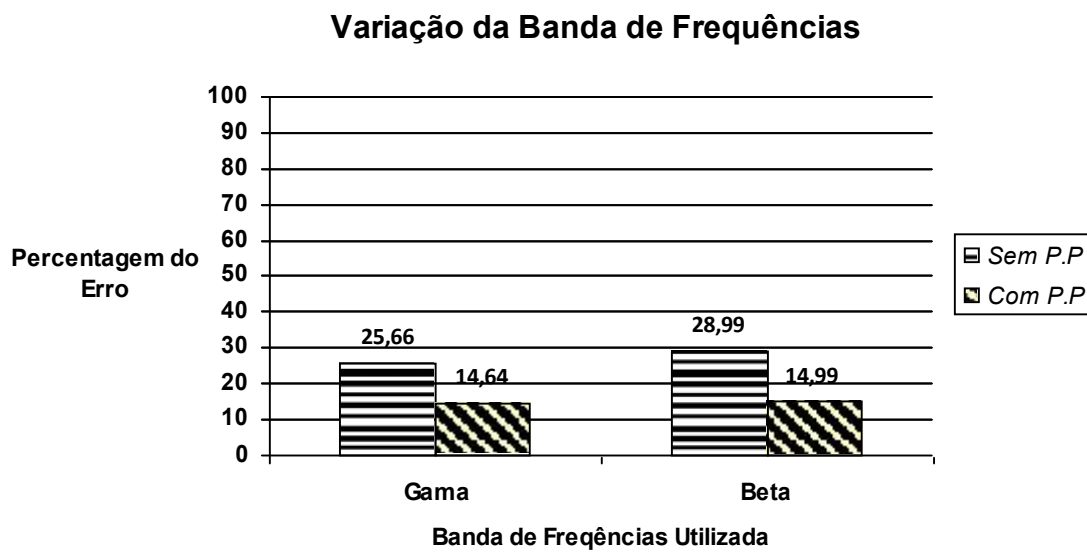


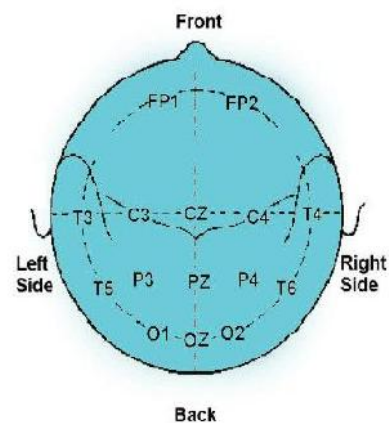
Fig.26 -> Variação da Banda de Frequências extraída

4.3.5 Redução Progressiva do Número de Canais Utilizados

- Testou-se a utilização de apenas 15 canais (excluíram-se os canais frontais)
- Testou-se a utilização de apenas 10 canais (excluíram-se os canais frontais, parietais e frontais-parietais)
- Testou-se a utilização de apenas 3 canais (apenas activos os canais occipitais)

Utilizando:

15 Canais activos
Filtro de ordem 4 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3



Obtém-se:

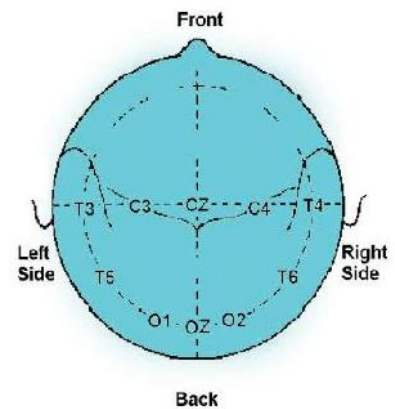
Percentagem de erro = **35,01%**

Percentagem de erro após Pós Processamento = **24,05%**

Fig.27 -> Resultados para a utilização de apenas 15 Canais activos

Utilizando:

10 Canais activos
Filtro de ordem 4 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3



Obtém-se:

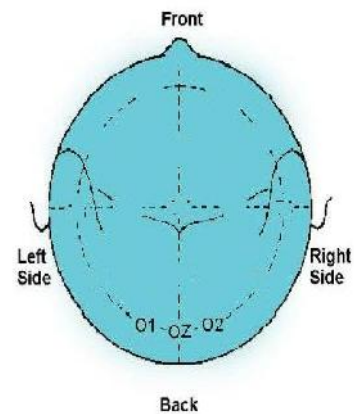
Percentagem de erro = **41,97%**

Percentagem de erro após Pós Processamento = **28,35%**

Fig.28 -> Resultados para a utilização de apenas 10 Canais activos

Utilizando:

3 Canais activos
Filtro de ordem 4 na banda de frequências gama
Todas as épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=1



Obtém-se:

Percentagem de erro = **78,45%**

Percentagem de erro após Pós Processamento = **73,51%**

Fig.29 -> Resultados para a utilização de apenas 3 Canais activos

Variação progressiva do número de Canais

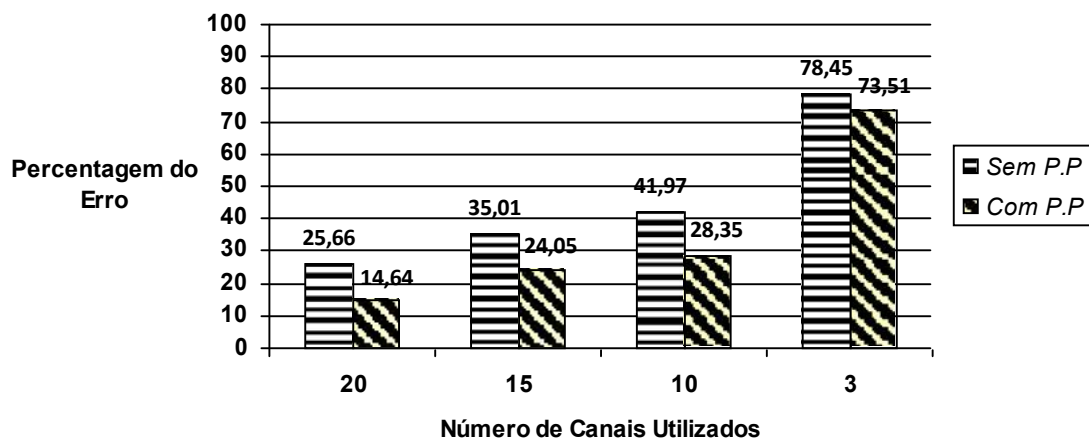


Fig.30 -> Redução progressiva do número de Canais utilizados

Analisando os resultados obtidos (figuras 27, 28 e 29) verifica-se que, ao reduzir-se o número de canais, o erro associado à classificação vai aumentando. Com o auxílio da fig.30 é possível fazer uma análise gráfica do aumento do erro de classificação à

medida que o número de canais é reduzido. Assim, para os pontos seguintes, recorreu-se sempre à utilização de 20 canais activos.

4.3.6 Redução Progressiva do Número de Épocas Utilizadas

- Foi testada a redução progressiva do número de épocas, utilizadas por pessoa, para as fases do treino e teste do classificador

Até aqui, foram constantemente utilizadas todas as épocas disponíveis para as fases do treino e teste na classificação, o que não se torna muito atractivo para uma posterior implementação uma vez que, por cada pessoa, tem-se mais de 200 épocas disponíveis.

Como se trabalha com uma frequência de amostragem de 250Hz e, como cada época é composta por 400 amostras, então cada uma tem um tempo de duração de 1,6s. Assim, e mesmo que consideremos apenas 200 épocas para treino e teste do classificador, a pessoa em questão teria de ser submetida a estímulos visuais durante mais de 5min, o que é certamente muito tempo para alguém ser autenticado.

Este ponto tem então como objectivo, verificar a viabilidade deste método para um número reduzido de épocas.

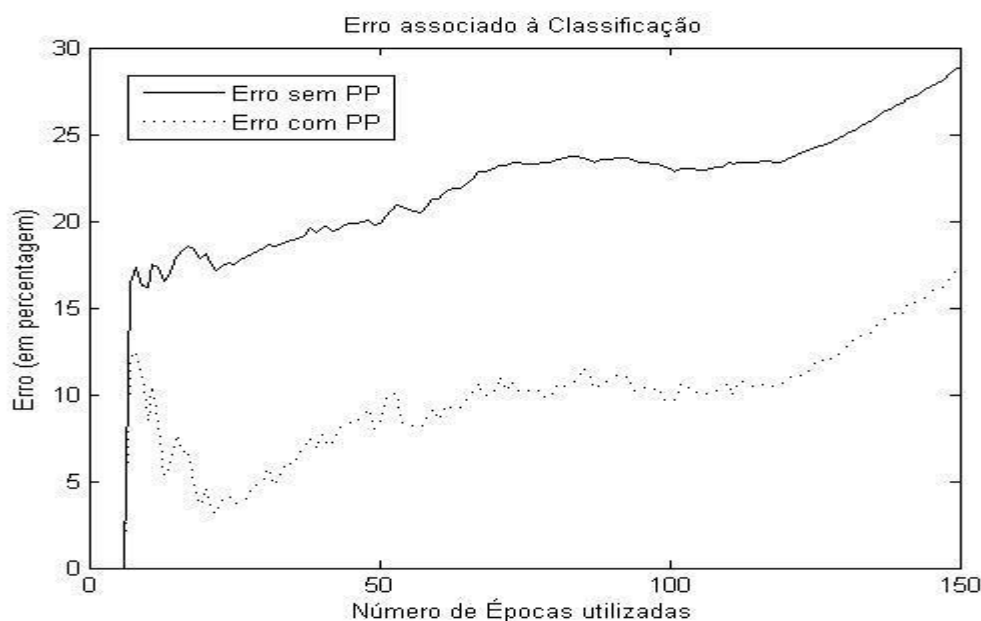


Fig.31 -> Erro associado à Classificação

Pela figura anterior (fig.31) constata-se que, não é por se utilizarem mais épocas para se efectuar a classificação que o erro de classificação diminui. Observa-se pelo contrário que um número reduzido de épocas permite obter melhores resultados.

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
120 Épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 23,53%
Percentagem de erro após Pós Processamento = 10,90%

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
80 Épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 23,46%
Percentagem de erro após Pós Processamento = 10,19%

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
40 Épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
K=3

Obtém-se:

Percentagem de erro = 19,62%
Percentagem de erro após Pós Processamento = 7,69%

Fig.32 -> Redução progressiva do número de Épocas utilizadas (utilização de 120, de 80 e de 40 Épocas)

Denota-se, pelos resultados apresentados na fig.31, que, todo este processo de classificação poderá ser, de facto, aplicado na realidade. Isto verifica-se visto que é possível observar que utilizando 20 a 30 épocas, por pessoa, para o processo de classificação permite otimizar os resultados, ou seja, obter o menor erro possível para o classificador. Isto torna-se muito vantajoso visto que utilizando apenas, por exemplo, 25 épocas para classificação, cada pessoa necessita apenas de 40s para ser autenticada, partindo do princípio que os seus dados existam no sistema.

Ainda sobre a fig.31, é de realçar, que o erro associado à classificação é zero para as 6 primeiras épocas porque a fase do Pós Processamento, como realiza sempre 5 iterações, necessita de pelo menos 7 épocas para começar a ser processada.

4.3.7 Alteração do Valor de K no Classificador

- Foram testados diferentes valores de K

Neste ponto, foram testados diferentes valores de K com o propósito de verificar se, de facto, o valor de $K=3$ utilizado até aqui, é realmente o mais adequado, ou seja, verificar se $K=3$ é o valor que induz menor erro na classificação. Para isso, partiu-se do exemplo anterior, cujos resultados são traduzidos graficamente pela fig.31 e, mantendo a mesma estrutura (as mesmas características), foi-se alterando o valor de K. As figuras seguintes (33, 34, 35, 36, 37, 38 e 39) ilustram os resultados obtidos.

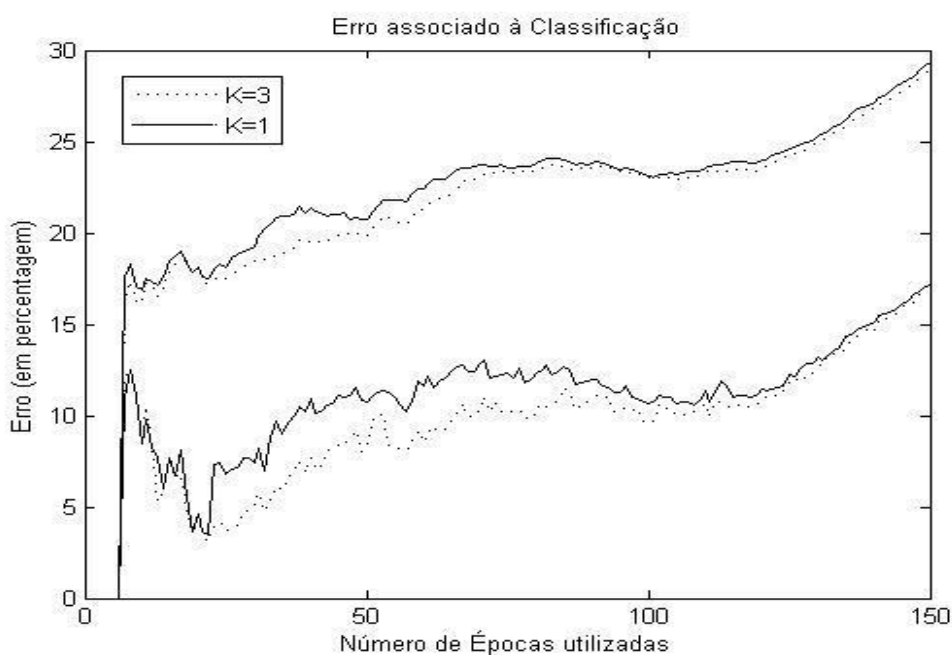


Fig.33 -> Erro associado à Classificação ($K=3$ Vs $K=1$)

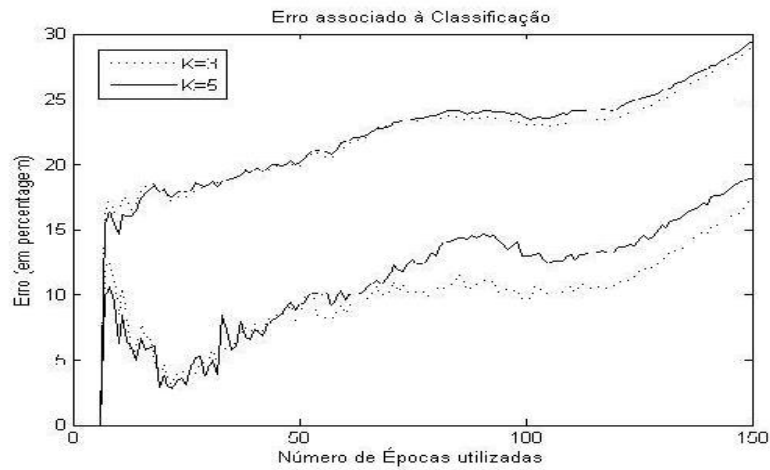


Fig.34 -> Erro associado à Classificação ($K=3$ Vs $K=5$)

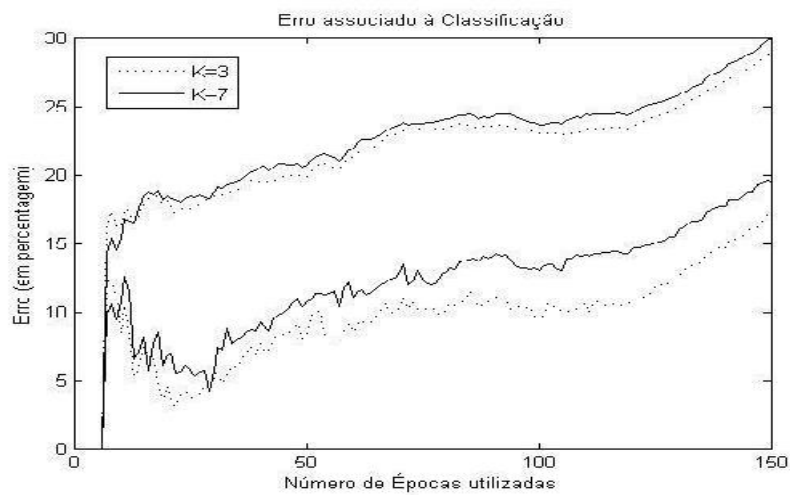


Fig.35-> Erro associado à Classificação ($K=3$ Vs $K=7$)

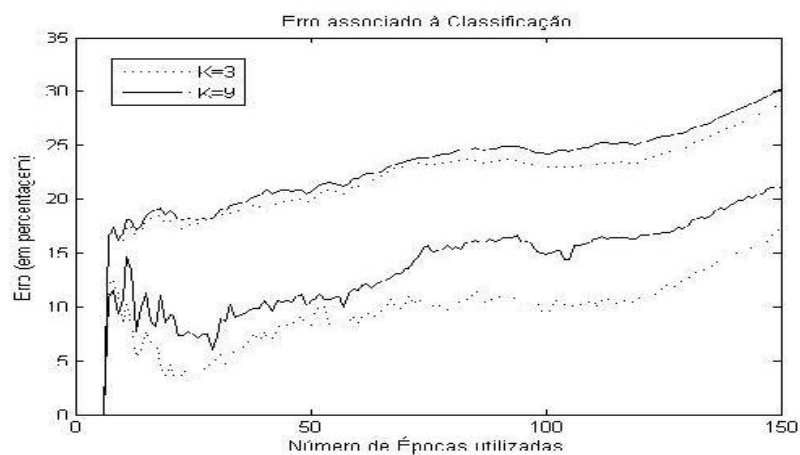


Fig.36 -> Erro associado à Classificação ($K=3$ Vs $K=9$)

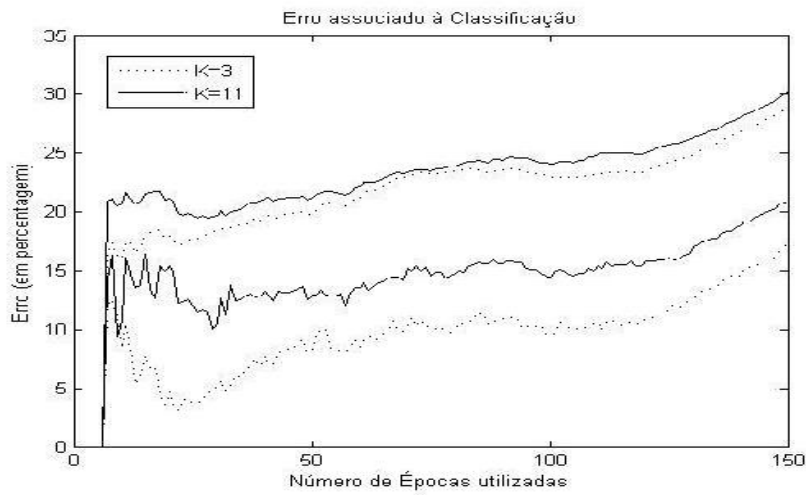


Fig.37 -> Erro associado à Classificação (K=3 Vs K=11)

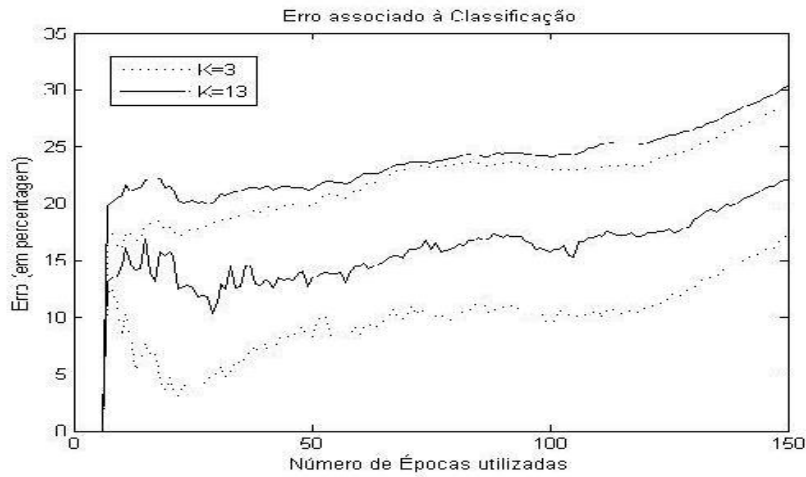


Fig.38 -> Erro associado à Classificação (K=3 Vs K=13)

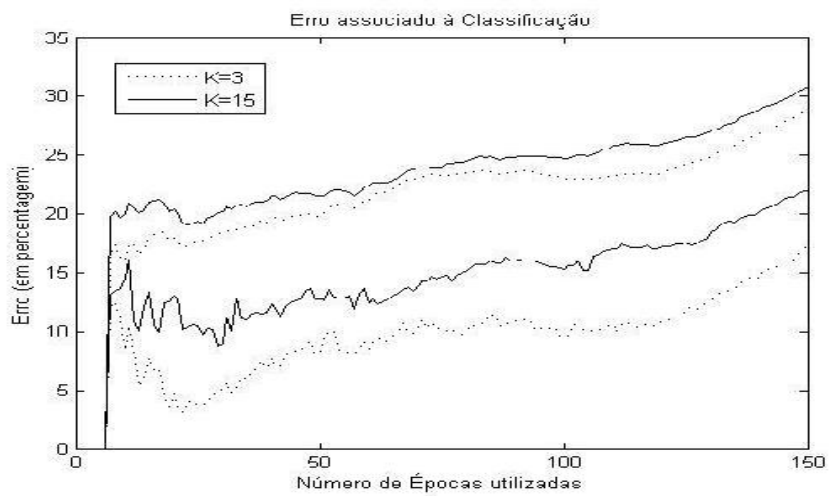


Fig.39 -> Erro associado à Classificação (K=3 Vs K=15)

Recorrendo às figuras anteriores para análise dos resultados, comprova-se que de facto $K=3$ é o melhor valor para K , tendo em vista obter o menor erro possível.

Ainda sobre as figuras anteriores (figuras 33, 34, 35, 36, 37, 38 e 39), é importante referir que traduzem os resultados com e sem Pós Processamento.

4.4 Outros Pontos a Considerar

4.4.1 Redução Progressiva do Número de Pessoas Utilizadas para Treino e Teste do Classificador

Até ao preciso momento, todos os resultados apresentados foram obtidos utilizando o conjunto total das 13 pessoas, tanto para o treino como para o teste do classificador.

Os resultados seguintes mostram o comportamento do classificador, quando se diminui progressivamente o conjunto de pessoas disponíveis tanto para treino como para teste do mesmo.

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
80 Épocas por pessoa
Treino do classificador com 13 pessoas
Teste do classificador com 13 pessoas
$K=3$

Obtém-se:

Percentagem de erro = 23,46%
Percentagem de erro após Pós Processamento = 10,19%

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
80 Épocas por pessoa
Treino do classificador com 11 pessoas
Teste do classificador com 11 pessoas
$K=3$

Obtém-se:

Percentagem de erro = 20,68%
Percentagem de erro após Pós Processamento = 10,11%

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
80 Épocas por pessoa
Treino do classificador com 8 pessoas
Teste do classificador com 8 pessoas
K=3

Obtém-se:

Percentagem de erro = 15,16%
Percentagem de erro após Pós Processamento = 7,34%

Utilizando:

20 Canais activos
Filtro de ordem 4 na banda de frequências gama
80 Épocas por pessoa
Treino do classificador com 6 pessoas
Teste do classificador com 6 pessoas
K=3

Obtém-se:

Percentagem de erro = 12,08%
Percentagem de erro após Pós Processamento = 9,79%

Fig.40 -> Resultados para a utilização de 13, de 11, de 8 e de 6 Pessoas tanto para Treino como para Teste do Classificador

Tal como se pode observar a partir dos resultados obtidos (fig.40 e fig.41), uma redução progressiva do número de pessoas utilizadas, implica também uma redução progressiva do erro de classificação (sem P.P.) pois o classificador tem menor margem de erro. Entre outros factores, isto deve-se ao facto de que, cada pessoa tem a si associado o seu próprio erro de classificação (erro de classificação individual).

Redução do Número de Pessoas

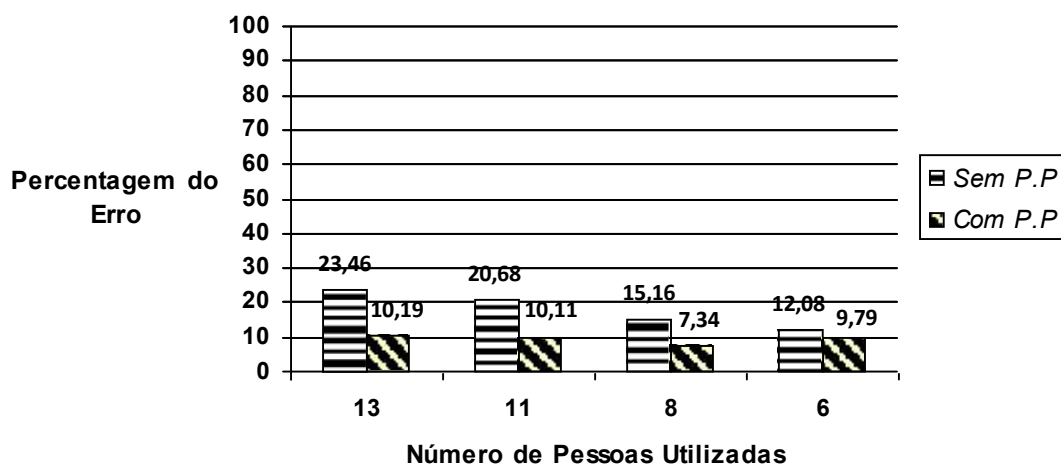


Fig.41 -> Redução progressiva do número de Pessoas utilizadas para Treino e Teste do Classificador

4.4.2 Erro de Classificação Individual

O erro de classificação individual, ou seja, o erro associado a cada pessoa, assume grande importância. No conjunto de todas as pessoas este, não só permite verificar quais são as principais responsáveis pelo erro associado ao classificador, como permite também evidenciar os benefícios do Pós Processamento, visto que, cada pessoa é analisada individualmente.

Utilizando-se o exemplo seguinte (fig.42) é possível observar como varia o erro de classificação de pessoa para pessoa.

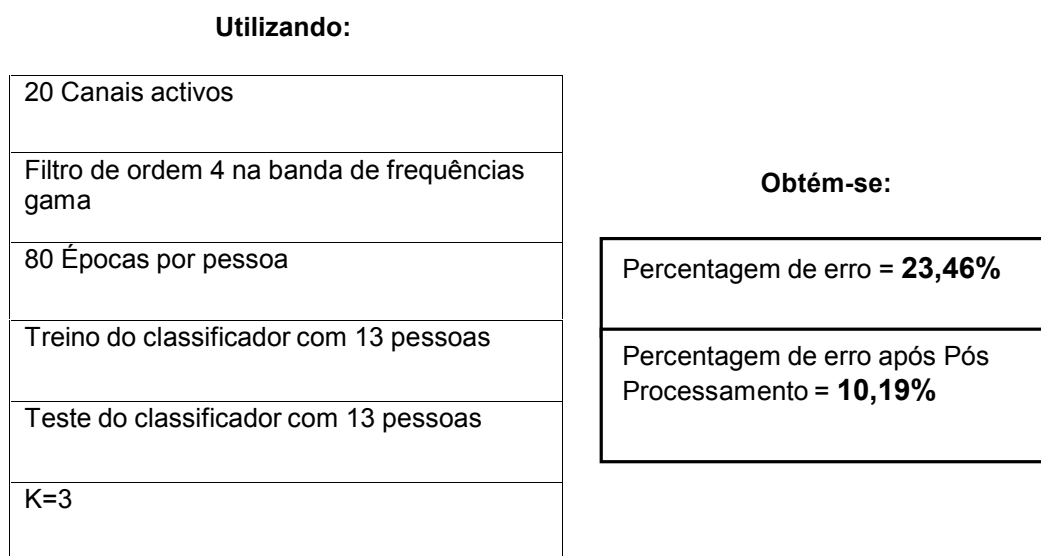


Fig.42 -> Exemplo escolhido para apresentar o Erro de Classificação Individual

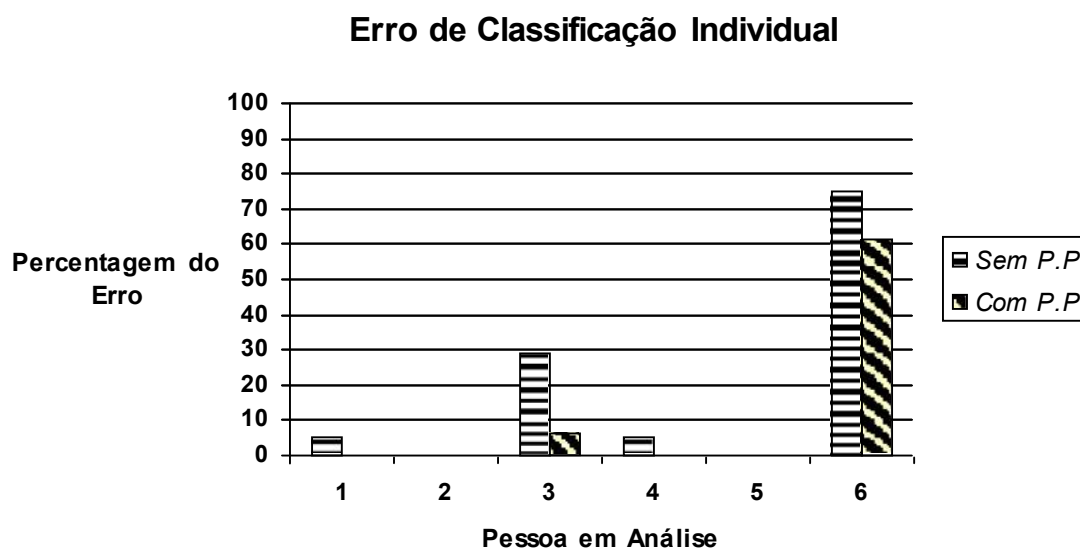


Fig.43 -> Erro de Classificação Individual (primeiras 6 Pessoas)

Erro de Classificação Individual (continuação)

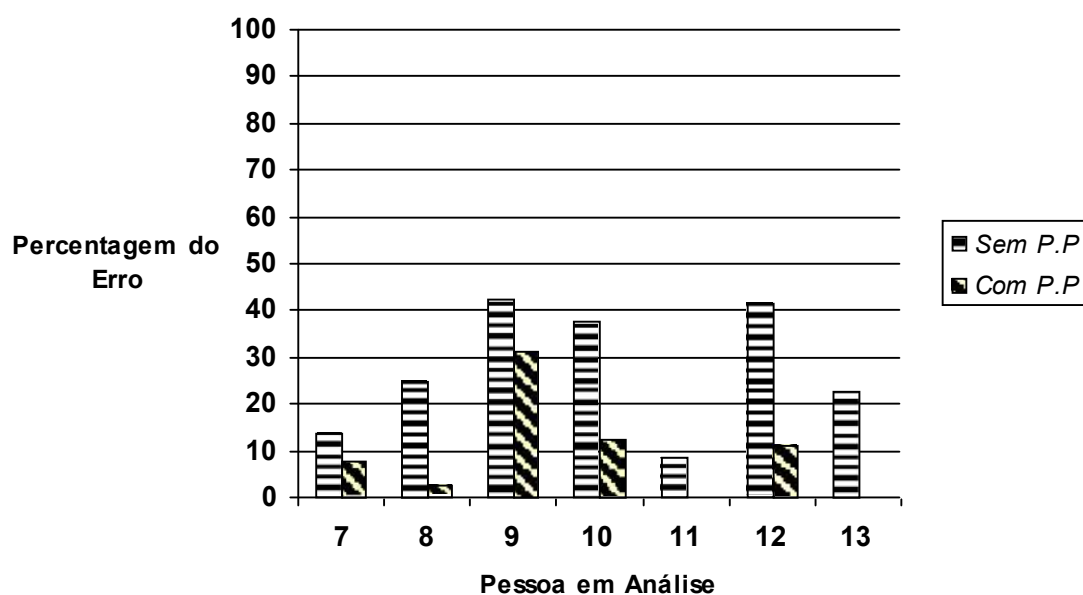


Fig.44 -> Erro de Classificação Individual (7 Pessoas seguintes)

Pessoa em Questão	Erro de Classificação
1	5,00%
2	0,00%
3	28,75%
4	5,00%
5	0,00%
6	75,00%
7	13,75%
8	25,00%
9	42,50%
10	37,50%
11	8,75%
12	41,25%
13	22,50%

Tabela 11 -> Erro de Classificação Individual sem PP

Pessoa em Questão	Erro de Classificação
1	0,00%
2	0,00%
3	6,25%
4	0,00%
5	0,00%
6	61,25%
7	7,50%
8	2,50%
9	31,25%
10	12,50%
11	0,00%
12	11,25%
13	0,00%

Tabela 12 -> Erro de Classificação Individual com PP

Os resultados obtidos e representados com auxílio da tabela 11 e da tabela 12 evidenciam que, de facto, existem pessoas com bons e pessoas com maus índices de classificação. As pessoas 6,9,10 e 12 são as pessoas que mais contribuem para o erro do classificador, como aliás se torna visível analisando as figuras 43 e 44.

De realçar também que, se forem somados todos os erros individuais e, esse valor dividido por 13 (número de pessoas), vai-se obter o erro total do classificador, como seria de esperar.

Capítulo 5

Conclusões

No capítulo anterior foram realizados vários testes, dos quais se obtiveram os resultados finais deste projecto.

Após a realização de todos esses testes, torna-se então possível chegar a uma conclusão, sobre as condições que melhor optimizam o processo de classificação.

Condições que optimizam o processo de classificação:

- Utilização de um filtro Butterworth de ordem 4 para filtrar o sinal EEG
- Utilização da banda de frequências gama como banda de frequências que se extrai do sinal EEG
- Utilização de 20 canais activos
- Utilização do valor 3 para a constante K do classificador
- Utilização de metade das épocas relativas aos estímulos das casas e metade das épocas relativas aos estímulos faciais para o treino e, à outra metade dos estímulos (casas e faces) para o teste do classificador
- Utilização de 20 a 30 épocas por pessoa para o processo de classificação
- Pós Processamento dos resultados obtidos

A utilização de todas estas condicionantes vai dar a origem a um erro (com P.P.) entre os 5,13% (30 épocas utilizadas) e os 3,15% (22 épocas utilizadas).

Trabalho Futuro

Este projecto poderá ter uma continuação no futuro, visto que, poderão ser realizados outros testes e utilizados outros processos, que não foram aqui implementados, de forma a obter um erro de classificação menor. Nesses teste e processos poderão incluir-se:

1. Utilização de maior número de eléctrodos
2. Utilização de uma base de dados mais vasta (maior número de pessoas)

3. Extracção de outra banda de frequências da experiência original
4. Utilização de outro tipo de classificador
5. Utilização de outro tipo técnica para o Pós Processamento dos resultados

Notas:

Nota1: Como já referido anteriormente, para obter a base de dados utilizada neste projecto, foi utilizado um filtro passa banda entre os 0.05 e os 30 Hz. Este ponto torna-se bastante importante visto que, no processamento do sinal EEG, recorreu-se a um filtro passa banda entre os 30 e os 50 Hz, ou seja, a parte do sinal que é filtrada neste projecto irá então apresentar uma energia reduzida.

Nota2: No que diz respeito ao Pós Processamento, esta fase permitiu diminuir, circunstancialmente, o erro de classificação. Não obstante, esta fase apresenta algumas características de relevo:

- São necessárias 7 épocas para se iniciar o Pós Processamento (utilizando 5 iterações)
- O erro pode aumentar de uma para outra iteração
- Tem melhor comportamento para classificações com erro reduzido do que para classificações com erro elevado
- O teste de um maior número de iterações permitiu concluir que, por cada uma, seria possível diminuir o erro em aproximadamente 0.5%. No entanto, por cada iteração irá ter-se um código mais extenso e, por isso, um processo mais demorado.

Torna-se também importante referir que, o Pós Processamento dos dados poderia ter sido realizado de outras formas distintas. Poderia, por exemplo, ser utilizada uma técnica por maioria, ou seja, analisando a sequência temporal de valores à saída do classificador, uma janela deslizante iria analisar os valores 3 a 3 ou 4 a 4 ou 5 a 5 (até um certo ponto), decidindo-se sempre pela maioria.

Nota3: Verifica-se, como já referido que, à medida que se reduz o número de épocas por pessoa, o erro de classificação diminui, sendo que, a utilização de 20 a 30 épocas por pessoa, traduz a obtenção do menor erro possível por parte do classificador. Uma possível explicação para esta ocorrência reside no facto de que, a partir de um certo número de potenciais evocados (entre 20 a 30) visualizados, o estado de espírito de uma pessoa se possa alterar, ou seja, a pessoa possa começar a mostrar sinais de cansaço. Este facto torna-se relevante, visto que, os mesmos resultados foram obtidos na tese realizada paralelamente a esta, referenciada em [R17] que utiliza um classificador diferente (SVM).

Nota4: Pode ser também utilizada uma diferente abordagem quanto à forma como se vê o erro de classificação. Esta, tem como base o erro de classificação individual. Nesta abordagem é definido um limite superior para o erro que, por norma, é 50%. Se o erro de classificação para determinada pessoa for inferior a 50% é sinal que o classificador vai autenticar correctamente a pessoa em questão e, se for superior a 50%, é sinal do contrário (má autenticação da pessoa por parte do classificador). Em suma, esta é uma forma diferente de olhar para a classificação, que torna este método de autenticação biométrica viável para uma futura utilização.

Bibliografia

- [1] *Beadvance*. (n.d.). Retrieved Julho 27, 2009, from Beadvance: http://www.beadvance.com/whats_br.asp
- [2] Magalhães, P. S. e Santos, H.D.: *Biometria e Autenticação*, Actas da 4ª conferência da Associação Portuguesa de Sistemas de Informação. Porto. Portugal. 15-17/10/2003 (edição em CD-ROM: ISBN 97 2-9354-42-1)
- [3] C.Gaio, A. d. (2009). *Biometria e seu impacto na sociedade*. Brasil.
- [4] Davies, S.: Touching Big Brother – How biometric technology will fuse flesh and machine, *Information Technology & People*, Vol 7, No. 4, 1994.
- [5] Santos, D. H. (2005). *Estudo dos padrões de Digitação e sua aplicação na autenticação biométrica*. Minho.
- [6] *Bioglobal*. (n.d.). Retrieved Agosto 13, 2009, from Bioglobal: <http://www.bioglobal.pt/content/index.php?action=detailfo&rec=164>
- [7] Prout, T. (2003). Trends in Personal Identification and Biometrics. *IDNewswire*, 2.
- [8] Liu, S. e Silverman, M.: A Practical Guide to Biometric Security Technology, IEEE Computer Society, www.computer.org/itpro/homepage/Jan_Feb/security3.htm, 2001
- [9] *CBA(Consultores Biométricos Associados)*. (n.d.). Retrieved Agosto 25, 2009, from CBA(Consultores Biométricos Associados): http://www.consultoresbiometricos.com.br/05_Bintroducao_definicao.php
- [10] *PsigWeb*. (n.d.). Retrieved Agosto 25, 2009, from PsigWeb: <http://virtualpsy.locaweb.com.br/index.php?art=356&sec=42>
- [11] Ebrahimi, T., & ET AL. (2003). Brain Computer Interface in Multimedia Communication. *IEEE Signal Processing Magazine*
- [12] Borges, D. F. (2005, Abril 29). A Neurofisiologia na Doença de Alzheimer.

- [13] Belur V. Dasarathy, editor (1991) *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, ISBN 0-8186-8930-7
- [14] Teknomo, B. K. (n.d.). *Kardi Teknomo's page*. Retrieved Outubro 8, 2009, from <http://people.revoledu.com/kardi/tutorial/KNN/index.html>
- [15] Isabel M. Santos, J. I. (2008). Differential effects of object-based attention on evoked potentials to fearful and disgusted faces.
- [16] Quintela, B. (2007). *Autenticação Biométrica através de actividade cerebral*. Aveiro.
- [17] Ferreira, A. (2009). *Sistema de Identificação pessoal baseado em sinais de EEG*. Aveiro.

Anexos

Para o processo de processamento do sinal recorreu-se às seguintes funções:

- fwplread
- rejeicao
- divisão_em_eventos
- filtro
- energia
- ident_total

Em seguida será exposto o código que compõe cada uma destas funções.

```
%% objectivo -> rejeição de épocas com artefactos
%%
%%[total,inrej]=fwplread(name1,name2)
% usage
% [t,v]=fwplread('nome.wpl') ou
% [t,v]=fwplread('nome.wpl', 'aaa')---dados em ficheiro mat... c,u,l
% Reading ficheiros *.wpl
```

```
function [total,inrej]=fwplread(name1,name2)
```

```
fid=fopen(name1); k=1;
[dummy,flag]=fread(fid,1,'char');
while flag
    c(k)=dummy;
    [l(k),flag]=fread(fid,1,'int');
    % Sem confirmação de boa leitura...
    [u(k),flag]=fread(fid,1,'int');
    % sem confirmação de boa leitura.....
    k=k+1; flag=0;
    [dummy,flag]=fread(fid,1,'char');
```

```
end
```

```
total=k-1;
```

```
inrej=find(c==3);
fclose(fid)
```

```
if nargin==2
    save(name2 , 'c', 'l', 'u');
end
```

-----//-----

```
%% objectivo -> leitura do sinal e rejeição das épocas
%%
%Leitura tem que ser feita pelo eeglab
%DATA -> é um ficheiro wpl por exemplo anabffa.wpl entre pelicas
%EEG -> corresponde a matrix 22x400x251, do sinal original do EEG
```

```
function [sinal,lenght]=rejeicao(DATA, EEG_s)
```

```

[total,inrej]=fwplread(DATA);    signal=EEG_s;    [a,b]=size(inrej);
lenght=total-b;    sinal=zeros(22,400,lenght);    k=1;    i=1;    j=1;
while j <= b
    if i == inrej(1,j)
        j=j+1;
        i=i+1;
    else
        sinal(:, :,k)=signal(:, :,i);
        k=k+1;
        i=i+1;
    end
end
for xx=i:total
    sinal(:, :,k)=signal(:, :,i);
    k=k+1;
    i=i+1;
end
end

```

-----//-----

```

%% objectivo -> fazer a divisão do sinal por evento
%%
%sinal -> provém da função rejeição
%Evento -> Evento(Época) que se pretende analisar

function[reconst] = divisao_em_eventos (sinal,Evento)

```

```

for i=1:20
    for j=1:400
        reconst(i,j)=sinal(i,j,Evento);
    end
end
end

```

-----//-----

```

%% Filtragem da banda gama (30-50Hz)
%%
%reconst -> é o sinal(de uma época) depois de reconstituído
%N -> é a ordem do filtro

```

```

function [filtrado]=Filtro(reconst,N)

x=reconst;
fs=250;    %frequencia de amostragem

W1=(30)/(125);    W2=(50)/(125);    W=[W1 W2];

[B,A] = BUTTER(N,W);
filtrado=(filtfilt(B,A,x'))';

```

```
%% Resposta do Filtro
```

```
% figure(1)  
% freqz(B,A,125,fs);
```

```
end
```

-----//-----

```
%% calculo da energia normalizada do sinal  
%%  
%como parametro de entrada temos o sinal (já filtrado) para uma época  
%devolve a energia dos 20 canais para uma época
```

```
function [energ]=energia(filtrado)
```

```
F=filtrado;  canais=1:1:20;  energ=sum(F'.^2);
```

```
norm=0;
```

```
for i=1:20  
    norm=energ(1,i)+norm;
```

```
end
```

```
energ=energ/norm; % normalização das energias
```

```
figure(1)  
canais=1:1:20;  
stem(canais,energ);
```

```
end
```

-----//-----

```
%% Processamento do sinal completo  
%%  
%DATA -> é o nome do ficheiro.wpl entre pelicas  
%EEG -> é o sinal com todas as épocas  
%N -> é a ordem do filtro butterworth
```

```
function [energ_t]=ident_total(DATA, EEG_s, N)
```

```
[sinal,lenght]=rejeicao(DATA, EEG_s);  
energ=zeros(lenght,20);
```

```
for i=1:lenght
```

```
    [reconst] = divisao_em_eventos (sinal,i);
```

```
    [filtrado]=Filtro(reconst,N);
```

```
    [energ]=energia(filtrado);
```

```
    energ_t(i,:)=energ;
```

```
    %organizar as energias de cada evento
```

```
    todas num vector (para cada pessoas)
```

```
end
```

-----//-----

Para se proceder à classificação foram utilizadas as seguintes funções:

- func_classknn
- func_classknn_mix
- func_classknn_alternado
- pos_proc_alterado
- erro_por_pessoa

De realçar que, como as funções func_classknn, func_classknn_mix e func_classknn_alternado apresentam bastantes semelhanças apenas será apresentado o código da função func_classknn.

Em seguida será exposto o código que compõe cada uma destas funções:

```
% funcao de classificacao pessoal usando o metodo do KNN
%Nesta função os estímulos referentes às fazes são usados para treino e os
%referentes às casas são usados para teste do classificador

%caract -> é o ficheiro .mat com todas as energias dos sujeitos
%n_pessoas -> é o numero de pessoas que se pretende usar para a classificação
%knn-> valor de K utilizado no classificador

function [erro_inicial,ypred,tst_necessario]=func_classknn(caract,n_pessoas,knn)

% if caract==1
%     load caracteristicas1;
% end

%% tamanho=zeros(1,26);
[tamanho(1,1),x]=size(energ_t_anabffa);
[tamanho(1,2),x]=size(energ_t_anabhfa);

[tamanho(1,3),x]=size(energ_t_anagffa);
[tamanho(1,4),x]=size(energ_t_anaghfa);

[tamanho(1,5),x]=size(energ_t_emmaffb);
[tamanho(1,6),x]=size(energ_t_emmahfb);

[tamanho(1,7),x]=size(energ_t_encaffa);
[tamanho(1,8),x]=size(energ_t_encahfa);

[tamanho(1,9),x]=size(energ_t_evalffa);
[tamanho(1,10),x]=size(energ_t_evalhfa);

[tamanho(1,11),x]=size(energ_t_guacffa);
[tamanho(1,12),x]=size(energ_t_guachfa);

[tamanho(1,13),x]=size( energ_t_laugffa);
[tamanho(1,14),x]=size( energ_t_laughfa);

[tamanho(1,15),x]=size(energ_t_loraffb);
[tamanho(1,16),x]=size(energ_t_lorahfb);

[tamanho(1,17),x]=size(energ_t_marcffb);
```

```

[tamanho(1,18),x]=size(energ_t_marchfb);

[tamanho(1,19),x]=size(energ_t_mcemffb);
[tamanho(1,20),x]=size(energ_t_mcemhfb);

[tamanho(1,21),x]=size(energ_t_noemffb);
[tamanho(1,22),x]=size(energ_t_noemhfb);

[tamanho(1,23),x]=size(energ_t_patvffa);
[tamanho(1,24),x]=size(energ_t_patvhfa);

[tamanho(1,25),x]=size(energ_t_yolbffa);
[tamanho(1,26),x]=size(energ_t_yolbhfa);

eventos=sum(tamanho);

%%
eventos_trn=0;
for i=1:2:26

    eventos_trn=eventos_trn+tamanho(1,i);
end
eventos_tst=0;
for i=2:2:26

    eventos_tst=eventos_tst+tamanho(1,i);
end

dados_tst=zeros(20,eventos_tst);
dados_trn=zeros(20,eventos_trn);

k=1;
for i=1:tamanho(1,1)
    dados_trn(:,k)=energ_t_anabffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,3)
    dados_trn(:,k)=energ_t_anagffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,5)
    dados_trn(:,k)=energ_t_emmaffb(i,:);
    k=k+1;
end

for i=1:tamanho(1,7)
    dados_trn(:,k)=energ_t_encaffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,9)
    dados_trn(:,k)=energ_t_evalffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,11)
    dados_trn(:,k)=energ_t_guacffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,13)
    dados_trn(:,k)=energ_t_laugffa(i,:);
    k=k+1;
end

for i=1:tamanho(1,15)
    dados_trn(:,k)=energ_t_loraffb(i,:);

```



```

        k=k+1;
    end

    for i=1:tamanho(1,17)
        dados_trn(:,k)=energ_t_marcffb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,19)
        dados_trn(:,k)=energ_t_mcemffb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,21)
        dados_trn(:,k)=energ_t_noemffb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,23)
        dados_trn(:,k)=energ_t_patvffa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,25)
        dados_trn(:,k)=energ_t_yolbffa(i,:);
        k=k+1;
    end

    %%
    k=1;

    for i=1:tamanho(1,2)
        dados_tst(:,k)=energ_t_anabhfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,4)
        dados_tst(:,k)=energ_t_anaghfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,6)
        dados_tst(:,k)=energ_t_emmahfb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,8)
        dados_tst(:,k)=energ_t_encahfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,10)
        dados_tst(:,k)=energ_t_evalhfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,12)
        dados_tst(:,k)=energ_t_guachfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,14)
        dados_tst(:,k)=energ_t_laughfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,16)
        dados_tst(:,k)=energ_t_lorahfb(i,:);

```

```

        k=k+1;
    end

    for i=1:tamanho(1,18)
        dados_tst(:,k)=energ_t_marchfb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,20)
        dados_tst(:,k)=energ_t_mcemhfb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,22)
        dados_tst(:,k)=energ_t_noemhfb(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,24)
        dados_tst(:,k)=energ_t_patvhfa(i,:);
        k=k+1;
    end

    for i=1:tamanho(1,26)
        dados_tst(:,k)=energ_t_yolbhfa(i,:);
        k=k+1;
    end

    %%
    eve_trn=0;
    for i=1:2:2*n_pessoas
        eve_trn=eve_trn+tamanho(1,i);
    end
    eve_tst=0;

    for i=2:2:2*n_pessoas
        eve_tst=eve_tst+tamanho(1,i);
    end

    ppl=zeros(20,eve_trn);

    k=1;
    for i=1:eve_trn
        ppl(:,i)=dados_trn(:,k);
        k=k+1;
    end

    trn.X=ppl;

    v=ones(1,eve_trn);      long=zeros(1,26);      k=1;

    for i=1:2:26
        long(1,k)=tamanho(1,i);
        k=k+1;
    end

    y=1;      i=1;      ev_count=0;

    while ev_count < eve_trn

        for k=1:long(1,i)
            v(1,y)=i*v(1,y);
            y=y+1;
        end

        ev_count=ev_count+long(1,i);
        i=i+1;
    end

```

```

trn.y=v;
trn.num_data=eve_trn;
trn.dim=2;
trn.name='estrutura_treino';

%%

ppl=zeros(20,eve_tst);          k=1;

for i=1:eve_tst
    ppl(:,i)=dados_tst(:,k);
    k=k+1;
end

tst.X=ppl;

v=ones(1,eve_tst);    long=zeros(1,26);    k=1;

for i=2:2:26
    long(1,k)=tamanho(1,i);
    k=k+1;
end

y=1;    i=1;    ev_count=0;

while ev_count < eve_tst

    for k=1:long(1,i)
        v(1,y)=i*v(1,y);
        y=y+1;
    end

    ev_count=ev_count+long(1,i);
    i=i+1;
end

tst.y=v;
tst.num_data=eve_tst;
tst.dim=2;
tst.name='estrutura_teste';

%%
model=knnrule(trn,knn) %k=1;
ypred=knnclass(tst.X,model);
erro_inicial=cerror(ypred,tst.y);
tst_necessario=tst.y;

end

-----//-----

%% Função responsável pelo pós processamento dos resultados
%ypred -> Sequência temporal de valores que resulta da classificação
%tst_necessario -> parâmetro da estrutura de teste do classificador que aqui é
necessário

%%

function[erro1,erro2,erro3,erro4,erro5,ypred1,ypred2,ypred3,ypred4,ypred5] =
pos_proc_alterado(ypred,tst_necessario)

tst.y=tst_necessario;

```

```

%% para eliminar uma irregularidade ex: 111211 -> 111111
%% 211121111211121

i=1;      j=2;      k=3;

for l=1:length(ypred)-2

if ypred(i)~=ypred(j)
    if ypred(j)~=ypred(k)
        if ypred(i)==ypred(k)

            ypred(j)=ypred(k);
        end
    end
end

i=i+1;
k=k+1;
j=j+1;
ypredl=ypred;

end
erro1=cerror(ypred,tst.y);

%% para duas irregularidades ex: 11122111 -> 11111111

i=1;      j=2;      k=3;      d=4;

for l=1:length(ypred)-3

if ypred(i)~=ypred(j)
    if ypred(j)~=ypred(k)
        if ypred(k)~=ypred(d)
            if ypred(i)==ypred(d)
                ypred(j)=ypred(i);
                ypred(k)=ypred(d);
            end
        end
    end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;

end

i=1;      j=2;      k=3;      d=4;

for l=1:length(ypred)-3

if ypred(i)~=ypred(j)
    if ypred(j)==ypred(k)
        if ypred(k)~=ypred(d)
            if ypred(i)==ypred(d)
                ypred(j)=ypred(i);
                ypred(k)=ypred(d);
            end
        end
    end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;

```

```

ypred2=ypred;

end
erro2=cerror(ypred,tst.y);

%% para 3 irregularidades ex: 11113331111 -> 11111111111

i=1;      j=2;      k=3;      d=4;      f=5;

for l=1:length(ypred)-4
    if ypred(i)~=ypred(j)
        if ypred(j)==ypred(k)
            if ypred(k)==ypred(d)
                if ypred(d)~=ypred(f)
                    if ypred(i)==ypred(f)
                        ypred(j)=ypred(i);
                        ypred(d)=ypred(f);
                        ypred(k)=ypred(j);
                    end
                end
            end
        end
    end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;
end

i=1;      j=2;      k=3;      d=4;      f=5;

for l=1:length(ypred)-4
    if ypred(i)~=ypred(j)
        if ypred(j)~=ypred(k)
            if ypred(k)==ypred(d)
                if ypred(d)~=ypred(f)
                    if ypred(i)==ypred(f)
                        ypred(j)=ypred(i);
                        ypred(d)=ypred(f);
                        ypred(k)=ypred(j);
                    end
                end
            end
        end
    end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;
end

i=1;      j=2;      k=3;      d=4;      f=5;

for l=1:length(ypred)-4
    if ypred(i)~=ypred(j)
        if ypred(j)==ypred(k)
            if ypred(k)~=ypred(d)
                if ypred(d)~=ypred(f)

```

```

        if ypred(i)==ypred(f)
            ypred(j)=ypred(i);
            ypred(d)=ypred(f);
            ypred(k)=ypred(j);
        end
    end
end

end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;
end

i=1;      j=2;      k=3;      d=4;      f=5;

for l=1:length(ypred)-4

if ypred(i)~=ypred(j)
    if ypred(j)~=ypred(k)
        if ypred(k)~=ypred(d)
            if ypred(d)~=ypred(f)
                if ypred(i)==ypred(f)
                    ypred(j)=ypred(i);
                    ypred(d)=ypred(f);
                    ypred(k)=ypred(j);
                end
            end
        end
    end
end

end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;

ypred3=ypred;

end
erro3=cerror(ypred,tst.y);

%% para 4 irregularidades ex: 111444411111 -> 111111111111

i=1;      j=2;      k=3;      d=4;      f=5;      g=6;

for l=1:length(ypred)-5

if ypred(i)~=ypred(j)
    if ypred(f)~=ypred(g)
        if ypred(i)==ypred(g)
            if ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f)
                ypred(j)=ypred(i);
                ypred(f)=ypred(g);
                ypred(k)=ypred(j);
                ypred(d)=ypred(f);

            elseif ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f)
                ypred(j)=ypred(i);
                ypred(f)=ypred(g);
                ypred(k)=ypred(j);
            end
        end
    end
end

```

```

        ypred(d)=ypred(f);

elseif ypred(j)~=ypred(k) || ypred(k)==ypred(d) || ypred(d)~=ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

elseif ypred(j)~=ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)==ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

elseif ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)~=ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

else ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f)
    ypred(j)=ypred(i);
    ypred(f)=ypred(g);
    ypred(k)=ypred(j);
    ypred(d)=ypred(f);

    end
end
end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;
g=g+1;

ypred4=ypred;

end
erro4=cerror(ypred,tst.y);

%% para 5 irregularidades ex: 1114444411111 -> 1111111111111

i=1;      j=2;      k=3;      d=4;      f=5;      g=6;      h=7;

for l=1:length(ypred)-6
    if ypred(i)~=ypred(j)
        if ypred(g)~=ypred(h)
            if ypred(i)==ypred(h)
                if ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)~=ypred(g)
                    ypred(j)=ypred(i);

```

```

        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)==ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)==ypred(f) ||
ypred(f)~=ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)~=ypred(d) || ypred(d)==ypred(f) ||
ypred(f)==ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)==ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)~=ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f) ||
ypred(f)~=ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)~=ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f) ||
ypred(f)==ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)~=ypred(g)
        ypred(j)=ypred(i);
        ypred(g)=ypred(h);
        ypred(k)=ypred(j);
        ypred(f)=ypred(g);
        ypred(d)=ypred(k);

```



```

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)==ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)==ypred(f) ||
ypred(f)~=ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)~=ypred(d) || ypred(d)==ypred(f) ||
ypred(f)==ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)~=ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)~=ypred(f) ||
ypred(f)==ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

elseif ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f) ||
ypred(f)~=ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);

else ypred(j)==ypred(k) || ypred(k)==ypred(d) || ypred(d)==ypred(f) ||
ypred(f)==ypred(g)
    ypred(j)=ypred(i);
    ypred(g)=ypred(h);
    ypred(k)=ypred(j);
    ypred(f)=ypred(g);
    ypred(d)=ypred(k);
end
end
end
end

i=i+1;
k=k+1;
j=j+1;
d=d+1;
f=f+1;
g=g+1;
h=h+1;

ypred5=ypred;

```

```
end
erro5=cerror(ypred,tst.y);
```

```
end
```

-----//-----

```
%% Função responsável por obter o erro para cada pessoa
```

```
%%
function [erro_pp] = erro_por_pessoa (ypred, eve_var, n_pessoas)
```

```
erro_pp=zeros(1,n_pessoas);      eve_ini=1;
```

```
for j=1:n_pessoas
```

```
count=0;
```

```
for i=eve_ini:eve_var*j
```

```
    if ypred(i)==j
        count=count+1;
```

```
    end
```

```
end
```

```
eve_err=eve_var-count;
```

```
erro_pp(j)=eve_err/eve_var;
```

```
eve_ini=eve_ini+eve_var;
```

```
end
```

```
end
```

-----//-----